

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 19:01:44 ; Search time 57.44 Seconds  
(without alignments)  
370.748 Million cell updates/sec

Title: US-09-495-823-7

Perfect score: 3012

Sequence: 1 MAPRCAGHPPPSPQACVC.....VPRYPKPRNRLNGGV 550

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1426.5	47.4	533	1 ARSB_HUMAN	P15848 homo sapien
2	1414.5	47.0	535	1 ARSB_FELCA	P33727 felis silve
3	1333	44.3	473	1 ARSB_RAT	P50430 rattus norv
4	864	28.7	285	1 ARSB_MOUSE	P50429 mus musculu
5	430.5	14.3	591	1 ARSF_HUMAN	P54793 homo sapien
6	429	14.2	535	1 ARS_PSEAE	P51691 pseudomonas
7	422.5	14.0	522	1 GA6S_HUMAN	P34059 homo sapien
8	416.5	13.8	567	1 ARS_STRPU	P50473 strongyloce
9	413	13.7	583	1 STS_HUMAN	P08842 homo sapien
10	407.5	13.5	589	1 ARSE_HUMAN	P15690 homo sapien
11	403	13.4	577	1 STS_RAT	P15589 rattus norv
12	399	13.2	593	1 ARSD_HUMAN	P51689 homo sapien
13	394	13.1	551	1 ASLA_ECOLI	P25549 escherichia
14	392.5	13.0	624	1 STS_MOUSE	P50427 mus musculu
15	378	12.5	507	1 ARSA_HUMAN	P15289 homo sapien
16	364	12.1	506	1 ARSA_MOUSE	P50428 mus musculu
17	356	11.8	464	1 ARS_KLEAE	P20713 klebsiella
18	349.5	11.6	551	1 ARS_HEMPU	P14000 hemocentrot
19	324	10.8	560	1 YDEN_ECOLI	P77318 escherichia
20	250	8.3	559	1 GL6S_CAPHI	P50426 capra hircu
21	248.5	8.3	552	1 GL6S_HUMAN	P15586 homo sapien
22	247	8.2	649	1 ARS_VOLCA	Q10723 volvox cart
23	241	8.0	497	1 YIDJ_ECOLI	P31447 escherichia
24	231	7.7	647	1 ARS_CHLRE	P14217 chlamydomon
25	214	7.1	502	1 SPHM_HUMAN	P51688 homo sapien
26	205.5	6.8	550	1 IDS_HUMAN	P22304 homo sapien
27	197	6.5	563	1 IDS_MOUSE	Q08890 mus musculu
28	184.5	6.1	512	1 BETC_RHIME	O69787 rhizobium m
29	135	4.5	647	1 YC46_HAEIN	P44135 haemophilus
30	122.5	4.1	638	1 YQGS_BACSU	P54496 bacillus su
31	116	3.9	553	1 HISS_EMENI	Q9p4P9 emericella
32	106.5	3.5	437	1 AM3E_ORYSA	P27934 oryza sativ
33	106.5	3.5	763	1 MDOB_ECOLI	P39401 escherichia

RESULT 1	ARSB_HUMAN	105	3.5	922	1	NRPI_RAT	Q9qwj9
ID	ARSB_HUMAN	103.5	3.4	1305	1	GAK_RAT	rattus norv
AC	P15848;	102.5	3.4	538	1	RKPI_RHIME	rattus norv
DT	01-APR-1990 (Rel. 14, Created)	102	3.4	700	1	CHAC_PEDHE	O52938 rhizobium m
DT	01-APR-1990 (Rel. 14, Last sequence update)	102	3.4	923	1	NRPI_MOUSE	O52988 pedobacter
DT	30-MAY-2000 (Rel. 39, Last annotation update)	102	3.4	1235	1	VGL2_CVMJH	P97333 mus musculu
DE	Arylsulfatase B precursor (EC 3.1.6.12) (ASB) (N-acetylgalactosamine-4-sulfatase) (G4S).	100.5	3.3	579	1	ASO_CUCMA	P11225 murine coro
GN	ARSB.	100	3.3	1066	1	HGPC_HAEIN	P24792 cucurbita m
OS	Homo sapiens (Human).	99	3.3	1132	1	JAK2_HUMAN	O9x442 haemophilus
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	98.5	3.3	2334	1	WAP2_BACSU	O60674 homo sapien
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	97.5	3.2	586	1	YEJW_SALTY	O07833 bacillus su
OX	NCBI_TaxID=9606;	97.5	3.2	1337	1	PTPJ_HUMAN	P40709 salmonella
[1]	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.						Q12913 homo sapien
RP	MEDLINE=90153994; PubMed=2303452;						
RA	Peters C., Schmidt B., Rommerskirch W., Rupp K., Zuehlendorf M.,						
RA	Vingron M., Meyer H.E., Pohlmann R., von Figura K.;						
RT	"Phylogenetic conservation of arylsulfatases. cDNA cloning and						
RT	expression of human arylsulfatase B.;"						
RL	J. Biol. Chem. 265:3374-3381(1990).						
[2]	SEQUENCE FROM N.A.						
RP	MEDLINE=90152677; PubMed=1968043;						
RA	Schuchman E.H., Jackson C.E., Desnick R.J.;						
RT	"Human arylsulfatase B: MOPAC cloning, nucleotide sequence of a full-						
RT	length cDNA, and regions of amino acid identity with arylsulfatases A						
and C.;"							
RL	Genomics 6:149-158(1990).						
[3]	SEQUENCE FROM N.A.						
RP	MEDLINE=93332648; PubMed=7687847;						
RA	Modarelli S., Rupp K., von Figura K., Peters C.;						
RT	"Structure of the human arylsulfatase B gene.;"						
RL	Biol. Chem. Hoppe-Seyler 374:327-335(1993).						
[4]	SEQUENCE OF 1-104 FROM N.A.						
RP	MEDLINE=92028992; PubMed=1930244;						
RA	Litjens T., Morris C.P., Gibson G.J., Beckmann K.R., Hopwood J.J.;						
RT	"Human N-acetylgalactosamine-4-sulphatase: protein maturation and						
RT	isolation of genomic clones.;"						
RL	Biochem. Int. 24:209-215(1991).						
[5]	2-AMINO-3-OXOPROPIONIC ACID MODIFICATION SITE.						
RP	MEDLINE=95354208; PubMed=7628016;						
RA	Schmidt B., Selmer T., Ingendoh A., von Figura K.;						
RT	"A novel amino acid modification in sulfatases that is defective in						
RT	multiple sulfatase deficiency.;"						
RL	Cell 82:271-278(1995).						
[6]	X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).						
RP	MEDLINE=97184692; PubMed=9032078;						
RA	Bond C.S., Clements P.R., Ashby S.J., Collyer C.A., Harrop S.J.,						
RA	Hopwood J.J., Guss J.M.;						
RT	"Structure of a human lysosomal sulfatase.;"						





```

ARB_RAT
ID ARSB_RAT STANDARD; PRT; 473 AA.
AC P50430;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Arylsulfatase B (EC 3.1.6.12) (ASB) (N-acetylgalactosamine-
4-sulfatase) (G4S) (Fragment).
GN ARSB.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Liver;
RX MEDLINE=96121368; PubMed=8575749;
RA Kunieda T.;
RT "Mucopolysaccharidosis type VI in rats: isolation of cDNAs encoding
arylsulfatase B, chromosomal localization of the gene, and
identification of the mutation.";
RL Genomics 29:582-587(1995).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 4-sulfate groups of the N-
acetyl-D-galactosamine 4-sulfate units of chondroitin sulfate and
dermatan sulfate.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Lysosomal.
CC -1- DISEASE: DEFECTS IN ARSB ARE THE CAUSE OF MAROTEAUX-LAMY
SYNDROME; ALSO KNOWN AS MUCOPOLYSACCHARIDOSIS TYPE VI (MPS-VI).
CC -1- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; D49434; BAA08412.1; -.
CC HSP; P15848; IFSU
DR InterPro: IPR000917; Sulfatase.
DR Pfam: PF00884; Sulfatase; 1.
DR PROSITE; PS00523; SULFATASE_1; 1.
DR PROSITE; PS00149; SULFATASE_2; FALSE NEG.
KW Hydrolase; Glycoprotein; Lysosome; Mucopolysaccharidosis.
FT NON_TER 1 1
FT MOD_RES 31 31 2-AMINO-3-OXOPROPIONIC ACID (BY
SIMILARITY).
FT ACT_SITE 87 87
FT DISULFID 57 461
FT DISULFID 61 95
FT DISULFID 121 132
FT DISULFID 345 387
FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 219 219 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 231 231 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 355 355 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 473 AA; 53320 MW; 4E114C923A24AF8F CRC64;

Query Match 44.3%; Score 1333; DB 1; Length 473;
Best Local Similarity 53.9%; Pred. No. 9.6e-98;
Matches 255; Conservative 70; Mismatches 122; Indels 26; Gaps 5;

QY 94 HGSBKPTTDLKLAAGVKKLNNYVQICTPSRSQFITGKYIHTGLQHSIIRTPQNCIL 153
Db 3 HGSVIRPHLDALAAGVLDNLYVQVLPCTPSRSQVLTGRTYQIHMGLQHVLTMCQNCV 62
QY 154 PLDNATLPQKLKEGVGTHVMGKWLGYRKECPMTRRGFDFTFGSLGSGDYTH---- 209

```

```

Db 63 PLDEKLLPQLLKDGASSTHMVGKWLHMYRKECLPTRRRGFDYTYFYLLGSEDTYTHEACA 122
QY 210 -YKCDSPCMGCGDLYENDNRAWDNGIYSTQMKTORVQQLASHNPKPFLFYATYQAV 268
Db 123 PIECLNGTRCALDRDGEPAKEYTD-IYSTNIFTKRTATLLIANHPPKPLFLYAFOSV 181
QY 269 HSPLOQAPGRYFEHYRSIINRRRYAAMLSCLEDAINNVTLALKTYGYNNSIIYSDDN 328
Db 182 HDPLQVPEYMEPYDFIODKHRIYAGVMVLLDEAVGNVTKALKSRGLWNVTLIFSDN 241
QY 329 GGQPTAGSNWPLRGSKGTWEGGIRAVGVFVHSPLLKNKGTVCKPEVHITOMYPTLISLA 388
Db 242 GGOIRSGNNWPLRGKGTWEGGIRGAGFVAPSLPKOKGVKSRMLHMTDPLTLVLA 301
QY 389 EGOIDEDIQLDGYDIWETISEGLRSPRVDIILHNTDPIY-----TRAKNGSWAAG 437
Db 302 GGSTHTGKPLDGFVWETISEGSPRVLELNLINDPFDGLPCPGKNTTPEKNDSPFLE 361
QY 438 YGIWNTATQSAIRVQHMKLLTGNPGYSDWPPQSFQSNLIG--PNRWNERITSSTGKSYWL 495
Db 362 HSAFNTSIHAGIRYKMKLLTGYPCGYWFPFPPSOSNISEVPS-----VDSPTKTLWL 414
QY 496 FNITADPYERVDLSNRTPGIVKLLRLRLSQNKTAVPVRYPPKDPKPRSNPLNG 548
Db 415 PDINRDPPEERHDVSRPHIVQNLRLSLQYYHEHSVPSYFPFPLDPRCDPKGTG 467

RESULT 4
ARB_MOUSE
ID ARSB_MOUSE STANDARD; PRT; 285 AA.
AC P50429;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Arylsulfatase B (EC 3.1.6.12) (ASB) (N-acetylgalactosamine-
4-sulfatase) (G4S) (Fragments).
GN ARSB OR AS1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE=92241876; PubMed=1572648;
RA Grompe M., Pieretti M., Caskey C.T., Ballabio A.;
RT "The sulfatase gene family: cross-species PCR cloning using the MOPAC
technique.";
RL Genomics 12:755-760(1992).
RN [2]
RP SEQUENCE OF 34-285 FROM N.A.
RX MEDLINE=96323200; PubMed=8710849;
RA Evers M., Saftig P., Schmidt P., Hafner A., McLochlin D.B.,
RA Schmahl W., Hess B., von Figura K., Peters C.W.B.;
RT "Targeted disruption of the arylsulfatase B gene results in mice
resembling the phenotype of mucopolysaccharidosis VI.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:8214-8219(1996).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 4-sulfate groups of the N-
acetyl-D-galactosamine 4-sulfate units of chondroitin sulfate and
dermatan sulfate.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Lysosomal.
CC -1- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; M82877; AAA37261.1; -.
CC EMBL; X92096; CAA63067.1; -.

```



```

DR HSSP; P15848; IFSU.
DR MGD; MGI:88075; Asi-s.
DR InterPro; IPR00917; Sulfatase.
DR Pfam; PF00884; Sulfatase; 1.
DR PROSITE; PS00523; SULFATASE_1; PARTIAL.
DR PROSITE; PS00149; SULFATASE_2; 1.
KW Hydrolase; Glycoprotein; Lysosome.
FT NON_TER 1
FT ACT_SITE 33 34
FT ACT_SITE 45 45
FT CARBOHYD 86 86
FT CARBOHYD 177 177
FT CARBOHYD 189 189
FT CARBOHYD 264 264
FT NON_TER 285 285
SQ SEQUENCE 285 AA; 31727 MW; 36BA5B98EA0770C9 CRC64;

Query Match 28.7%; Score 864; DB 1; Length 285;
Best Local Similarity 49.8%; Pred. No. 5.3e-61;
Matches 164; Conservative 45; Mismatches 68; Indels 52; Gaps 3;

QY 91 VGHGSEIKTTLDKLAEGVKLENNYVQPICTPSRSQFITGKYIHTGLOHSIIRPTQP 150
Db 4 LGFHGSVIRTPHLDALAAGVVLNYVQP----- 33

QY 151 NCLPLDNATLPQKLKEVGYSTHMYGKWHLGFYRKECMPTRRGFDTFGSLGSGDYTHY 210
Db 34 -----AGYATHMYGKWHLGMNRYKECLPTRRGFDTFYGLGSEDTYTHE 77

QY 211 KCD-----SPGMGYDLYENDNAWDYNGIYSTOMYTORVQOILASHNPTRKPIFLYIAY 265
Db 78 ACAPTESLNGTRCALDLRDGEAPKEYNN-IYSTNIETKRAITVIANHPPEKPLFLYLA 136

QY 266 QAVHSPLQAPGRYEHYESIINIRRRYAAMLSCLEAINNVTALKTYGYNNIIYS 325
Db 137 QSVHDPLQAPPEYEPYGFIDOKHRIIRYAGVSLMDAEAVGNVTALKSHGLNNTVFLFS 196

QY 326 SDNGQPTAGSNPLRSGKTYEGGIRAVGFVHSPILKKNKGVCKRPEVHTDWTPLI 385
Db 197 TDNGQTRSGGNPLRGRKGTLEGGIRGTGFVASPLLKQKGVKSRMLHISDLPLTV 256

QY 386 SLAGQDEIDQLDGYDIWETISEGLRSP 414
Db 257 DLGGSTNGTKPLDGFNNMKTISEGHPSP 285

RESULT 5
ARSF_HUMAN STANDARD; PRT; 591 AA.
AC P54793;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Arylsulfatase f precursor (EC 3.1.6.-) (ASF).
GN ARSF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=97336043; PubMed=912838;
RA Puca A.A., Zollo M., Repetto M., Andolfi G., Guffanti A., Simon G.,
RA Ballabio A., Franco B.;
RT "Identification by shotgun sequencing, genomic organization, and
RT functional analysis of a fourth arylsulfatase gene (ARSF) from the
RT xp22.3 region."
RL Genomics 42:192-199(1997).
RN [2]
RP SEQUENCE OF 370-423 FROM N.A.
RC TISSUE=Kidney;

```

```

RX MEDLINE=95236447; PubMed=7720070;
RA Franco B., Meroni G., Parenti G., Levilliers J., Bernard L.,
RA Gebbia M., Cox L., Maroteaux P., Sheffield L., Rappold G.A.,
RA Andria G., Petit C., Ballabio A.;
RT "A cluster of sulfatase genes on xp22.3: mutations in
RT chondrodysplasia punctata (CDPX) and implications for warfarin
RT embryopathy."
RL Cell 81:15-25(1995).
CC -|- ENZYME REGULATION: NOT INHIBITED BY DHEAS OR WARFARIN.
CC -|- MITOCHONDRIAL: OPTIMUM PH IS 8.
CC -|- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X97868; CAA66462.1; -
DR HSSP; P15848; IFSU.
DR MW; 300003; -
DR InterPro; IPR00917; Sulfatase.
DR Pfam; PF00884; Sulfatase; 1.
DR PROSITE; PS00523; SULFATASE_1; 1.
DR PROSITE; PS00149; SULFATASE_2; 1.
KW Hydrolase; Glycoprotein; Signal.
FT SIGNAL 1 19
FT CHAIN 20 591
FT MOD_RES 79 79
FT ACT_SITE 140 140
FT ACT_SITE 51 51
FT CARBOHYD 118 118
FT CARBOHYD 337 337
FT CARBOHYD 337 337
SQ SEQUENCE 591 AA; 66003 MW; 85B4B9BF3803A771 CRC64;

Query Match 14.3%; Score 430.5; DB 1; Length 591;
Best Local Similarity 25.1%; Pred. No. 2.5e-26;
Matches 143; Conservative 86; Mismatches 189; Indels 151; Gaps 21;

QY 73 TSQPHILFADQDQGRDVGHYGSE- IKPTDLDKLAAGVKL-ENYVQPICTPSRSQFI 130
Db 28 TTRPNIVLIMVDDLIGDLGCGYNDTMRTPHIDRLAREGVRLTQHISAASLCSPSRAFL 87

QY 131 TGKYOIHTGLOHS-----IIRPTQPNCLPLDNATLPQKLKEVGYSTHMYGKWHLGF---- 181
Db 88 TGRYPPIRSGMVSSGNRRRVIQNLAVPAGLPLNETTLAALLKKGYSTGLIGKWHQGLNCD 147

QY 182 YRKEC-MPTRRGFDTFGGS----- 199
Db 148 RSDQCHHPYNYGFDYIYGMPTLVDSQWPDSPRNTLAFESQMLCVOLVAIAIULTLTFG 207

QY 200 -----LLAGSDYYTHYKCDSPGCMGYDLYENDNAWD-----YDN 234
Db 208 KLSGWSVPWLLIFSMILFIFLLGVAPSSH---TSP-----LY-----WDCLLMRGH 252

QY 235 GIYSTQMYTORVQOILA-----SHNPTKPIFLYIAVQVHSPLOAPGRFEHYRSLIN 287
Db 253 EITEQPMKAERAGSTMVKEATISFLERHSKETFLFFSFHLVHTPLTPTDDETGTSKHL- 311

QY 288 INRRRYAAMLSCLEAINNVTALKTYGYNNIIYSDDNGQPTA-----GGSNWP 340
Db 312 -----YGNVEMDSMWGKILDAIDDFGLRNTLVYFTSDHGHLEARRGHAQLGNGI 366

QY 341 LRSGKGT-YWEGGIRAVGFVHSPILKKNKGVCKRPEVHTDWTPLTSLIAEGDIDOLD 399
Db 367 YKGGKMGWEGGIRVPGIVRWPKVPAGRLIKEPTSLMDILPTVASVSGSLPDQVID 426

QY 400 GYDIWETISEGLRSPRVLDLHNDIPYTKA-----KNGSMAAGYGIWNTATQSAIRVOHW 454

```



RA Tomatsu S., Fukuda S., Masue M., Sukegawa K., Fukao T., Yamagishi A.,  
RA Hori T., Iwata H., Ogawa T., Nakashima Y., Hanyu Y., Hashimoto T.,  
RA Titani K., Oyama R., Suzuki M., Yagi K., Hayashi Y., Orii T.;  
RT "Morquio A disease: isolation, characterization and expression of full-  
RL length cDNA for human N-acetylgalactosamine-6-sulfate sulfatase.";  
RN Biochem. Biophys. Res. Commun. 181:677-683(1991).  
RP [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-95095267; PubMed-8001980;  
RA Morris C.P., Guo X.H., Apostolou S., Hopwood J.J., Scott H.S.;  
RT "Morquio A syndrome: cloning, sequence, and structure of the human N-  
RT acetylgalactosamine 6-sulfatase (GALNS) gene.";  
RL Genomics 22:652-654(1994).  
RN [3]  
RP VARIANT MPS-IVA LYS-204.  
RX MEDLINE-92395122; PubMed-1522213;  
RA Fukuda S., Tomatsu S., Masue M., Sukegawa K., Iwata H., Ogawa T.,  
RA Nakashima Y., Hori T., Yamagishi A., Hanyu Y., Morooka K., Kitan T.,  
RA Hashimoto T., Orii T.;  
RT "Mucopolysaccharidosis type IVA. N-acetylgalactosamine-6-sulfate  
RT sulfatase exonic point mutations in classical Morquio and mild  
RT cases.";  
RL J. Clin. Invest. 90:1049-1053(1992).  
RN [4]  
RP VARIANTS MPS-IVA.  
RX MEDLINE-95397840; PubMed-7668283;  
RA Tomatsu S., Fukuda S., Cooper A., Wraith J.E., Maruf Rezvi G.,  
RA Yamagishi A., Yamada N., Kato Z., Isogai K., Sukegawa K., Kondo N.,  
RA Suzuki Y., Shinozawa N., Orii T.;  
RT "Mucopolysaccharidosis IVA: Identification of a common missense  
RT mutation III13F in the N-Acetylgalactosamine-6-sulfate sulfatase  
RT gene.";  
RL Am. J. Hum. Genet. 57:556-563(1995).  
RN [5]  
RP VARIANTS MPS-IVA.  
RX MEDLINE-95315929; PubMed-7795586;  
RA Ogawa T., Tomatsu S., Fukuda S., Yamagishi A., Maruf Rezvi G.,  
RA Sukegawa K., Kondo N., Suzuki Y., Shinozawa N., Orii T.;  
RT "Mucopolysaccharidosis IVA: screening and identification of mutations  
RT of the N-acetylgalactosamine-6-sulfate sulfatase gene.";  
RL Hum. Mol. Genet. 4:341-349(1995).  
RN [6]  
RP VARIANTS MPS-IVA ARG-77; TRP-90; VAL-96; LEU-151; GLY-230 AND THR-291.  
RX MEDLINE-95359983; PubMed-7633425;  
RA Tomatsu S., Fukuda S., Cooper A., Wraith J.E., Maruf Rezvi G.,  
RA Yamagishi A., Yamada N., Kato Z., Isogai K., Sukegawa K., Kondo N.,  
RA Suzuki Y., Shinozawa N., Orii T.;  
RT "Mucopolysaccharidosis type IVA: identification of six novel  
RT mutations among non-Japanese patients.";  
RL Hum. Mol. Genet. 4:741-743(1995).  
RN [7]  
RP VARIANT MPS-IVA SER-487.  
RX MEDLINE-96047158; PubMed-7581409;  
RA Tomatsu S., Fukuda S., Cooper A., Wraith J.E., Yamada N., Isogai K.,  
RA Kato Z., Sukegawa K., Kondo N., Suzuki Y., Shinozawa N., Orii T.;  
RT "Two new mutations, Q473X and N487S, in a Caucasian patient with  
RT mucopolysaccharidosis IVA (Morquio disease).";  
RL Hum. Mutat. 6:195-196(1995).  
RN [8]  
RP VARIANTS MPS-IVA ALA-138; SER-151 AND LEU-151.  
RX MEDLINE-96216146; PubMed-8651279;  
RA Tomatsu S., Fukuda S., Yamagishi A., Cooper A., Wraith J.E., Hori T.,  
RA Kato Z., Yamada N., Isogai K., Sukegawa K., Kondo N., Suzuki Y.,  
RA Shinozawa N., Orii T.;  
RT "Mucopolysaccharidosis IVA: four new exonic mutations in patients  
RT with N-acetylgalactosamine-6-sulfate sulfatase deficiency.";  
RL Am. J. Hum. Genet. 58:950-962(1996).  
RN [9]  
RP VARIANTS MPS-IVA CYS-94 AND VAL-97.  
RX MEDLINE-96423834; PubMed-8826435;  
RA Cole D.E.C., Fukuda S., Gordon B.A., Rip J.W., Lecouteur A.N.,  
RA Rupar C.A., Tomatsu S., Ogawa T., Sukegawa K., Orii T.;  
RT "Heteroallelic missense mutations of the galactosamine-6-sulfate

RT sulfatase (GALNS) gene in a mild form of Morquio disease (MPS IVA).";  
RL Am. J. Med. Genet. 63:558-565(1996).  
RN [10]  
RP VARIANTS MPS-IVA.  
RX MEDLINE-97442274; PubMed-9298823;  
RA Bunge S., Kleijer W.J., Tylki-Szymanska A., Steglich C., Beck M.,  
RA Tomatsu S., Fukuda S., Poorthuis B.J.H.M., Czartoryska B., Orii T.,  
RA Gal A.;  
RT "Identification of 31 novel mutations in the N-acetylgalactosamine-6-  
RT sulfatase gene reveals excessive allelic heterogeneity among patients  
RT with Morquio A syndrome.";  
RL Hum. Mutat. 10:223-232(1997).  
RN [11]  
RP VARIANTS MPS-IVA.  
RX MEDLINE-98041700; PubMed-9375852;  
RA Tomatsu S., Fukuda S., Cooper A., Wraith J.E., Ferreira P.,  
RA di Natale P., Tortora P., Fujimoto A., Kato Z., Yamada N., Isogai K.,  
RA Yamagishi A., Sukegawa K., Suzuki Y., Shinozawa N., Kondo N.,  
RA Sly W.S., Orii T.;  
RT "Fourteen novel mucopolysaccharidosis IVA producing mutations in  
RT GALNS gene.";  
RL Hum. Mutat. 10:368-375(1997).  
RN [12]  
RP VARIANTS MPS-IVA.  
RX MEDLINE-98180718; PubMed-9521421;  
RA Yamada N., Fukuda S., Tomatsu S., Muller V., Hopwood J.J., Nelson J.,  
RA Kato Z., Yamagishi A., Sukegawa K., Kondo N., Orii T.;  
RT "Molecular heterogeneity in mucopolysaccharidosis IVA in Australia  
RT and Northern Ireland: nine novel mutations including T312S, a common  
RT allele that confers a mild phenotype.";  
RL Hum. Mutat. 11:202-208(1998).  
RN [13]  
RP VARIANTS SER-393 AND MET-488.  
RX MEDLINE-98112415; PubMed-9452036;  
RA Tomatsu S., Fukuda S., Cooper A., Wraith J.E., Yamagishi A., Kato Z.,  
RA Yamada N., Isogai K., Sukegawa K., Suzuki Y., Shinozawa N., Kondo N.,  
RA Orii T.;  
RT "Fifteen polymorphisms in the N-acetylgalactosamine-6-sulfate  
RT sulfatase (GALNS) gene: diagnostic implications in Morquio disease.";  
RL Hum. Mutat. Suppl. 1:S42-S46(1998).  
RN -I- CATALYTIC ACTIVITY: Hydrolysis of the 6-sulfate groups of the N-  
CC acetyl-D-galactosamine 6-sulfate units of chondroitin sulfate and  
CC of the D-galactose 6-sulfate units of keratan sulfate.  
CC -I- SUBUNIT: OLIGOMER OF DISULFIDE LINKED 40- AND 15-KDA POLYPEPTIDES.  
CC -I- SUBCELLULAR LOCATION: Lysosomal.  
CC -I- DISEASE: Defects in GALNS are the cause of mucopolysaccharidosis  
CC - type IVA (MPS-IVA) (also known as Morquio A syndrome) which is  
CC characterized by specific spondyloepiphyseal dysplasia, short  
CC trunk dwarfism, coxa valga, odontoid hypoplasia, corneal  
CC opacities, preservation of intelligence, and excessive urinary  
CC excretion of keratan sulfate and chondroitin-6-sulfate. Severely  
CC affected patients usually die of cardiopulmonary disturbance or  
CC cervical cord compression in the second or third decade of life.  
CC -I- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; D17629; BAA04535.1;  
DR EMBL; D17616; BAA04535.1; JOINED.  
DR EMBL; D17617; BAA04535.1; JOINED.  
DR EMBL; D17618; BAA04535.1; JOINED.  
DR EMBL; D17619; BAA04535.1; JOINED.  
DR EMBL; D17620; BAA04535.1; JOINED.  
DR EMBL; D17621; BAA04535.1; JOINED.  
DR EMBL; D17622; BAA04535.1; JOINED.  
DR EMBL; D17623; BAA04535.1; JOINED.  
DR EMBL; D17624; BAA04535.1; JOINED.



```
QY 116 YVQPICTPSRSQBITKYOIHITGL--OHSIIRPTQPCNCLPLDNATLPQKLKEVGYSTHM 173
| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 109 YGSDVCTPSRSIAITGPIRTGYVEIRIFPWTITGLPLVEVTAIAEMKAGAGYTTGM 168
| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 174 VGKWHLCFYRKEC-----MPTRGDFDFGSLGSGDYTHYKCDSPGCMGYDLYENDNA 228
| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 169 VGKWHLGINSSSDGAHLFANKGFD-FVGHNLPPFGN---SWRCDDTGL--HODFPPTNA 222
| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 229 AMDYDNGIYQMTYQR-VOQIL-----ASHNPTKPIFLYIAYQAVHSPLOAPGRYFE 280
| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 223 CFLYNSTVAQPQKHGLTQLLRDDTVGVFTEDNVNPKPFMYVFAHHTSL-----FS 276
| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 281 HYRSIINRRYAAULSCIDEAINNVTLALKTYGFYNNIIYSSDNG-----GOPTA 334
| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 277 SDFSCSTRRGYDGNLREMDQAEIQVITLVNDIDDDNTVIFTSDBGPHREYCGE--- 333
| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 335 GGSNWPLRGSGTGWEGIRAVGFVHSPLLKNGKTCVKEPHVITDWPYTLISLAEGOIDE 394
| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 334 GGDANVFRGKGSGWEGGHRPIYIVYWPGTISPQ-VSHEIVTSMDIATAVNLGGSQLPT 392
| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 395 DIQLDGDYDIWETISEGLRSPRVDILHNDIPYTKAKNGSWAAQY 438
| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 393 DRIYDGGKLSVLEGASSPHDFFYCKDTLMAVRVGYKAHF 436
| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
RESULT 9
STS_HUMAN STANDARD; PRT; 583 AA.
AC P08842;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Steryl-sulfatase precursor (EC 3.1.6.2) (Steroid sulfatase) (Steryl-
DE sulfate sulfohydrolase) (Arylsulfatase C) (ASC).
GN STS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89340479; PubMed=2668275;
RA Stein C., Hille A., Seidel J., Rijnbout S., Waheed A., Schmidt B.,
RA Geuze H., von Figura K.;
RT "Cloning and expression of human steroid-sulfatase. Membrane
RT topology, glycosylation, and subcellular distribution in BHK-21
RT cells.";
RL J. Biol. Chem. 264:13865-13872(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=8717642; PubMed=3032454;
RA Yen P.H., Allen E., Marsh B., Mohandas T., Wang N., Taggart R.T.,
RA Shapiro L.J.;
RT "Cloning and expression of steroid sulfatase cDNA and the frequent
RT occurrence of deletions in STS deficiency: implications for X-Y
RT interchange.";
RL Cell 49:443-454(1987).
RN [3]
RP SEQUENCE OF 134-17A AND 461-583 FROM N.A.
RX MEDLINE=89077541; PubMed=3203382;
RA Yen P.H., Marsh B., Allen E., Tsai S.P., Ellison J., Connolly L.,
RA Neiswander K., Shapiro L.J.;
RT "The human X-linked steroid sulfatase gene and a Y-encoded
RT pseudogene: evidence for an inversion of the Y chromosome during
RT primate evolution.";
RL Cell 55:1123-1135(1988).
RN [4]
RP SEQUENCE OF 22-45.
RC TISSUE=Liver;
RX MEDLINE=89352671; PubMed=2765556;
RA Kawano J.-I., Kotani T., Ohtaki S., Minamino N., Matsuo H., Oinuma T.,
RA Akawa E.;
RT "Characterization of rat and human steroid sulfatases.";
```

```
RL Biochim. Biophys. Acta 997:199-205(1989).
RN [5]
RP VARIANTS XLI LEU-341; ARG-372 AND TYR-446.
RX MEDLINE=92170784; PubMed=1539590;
RA Basler E., Gronpe M., Parenti G., Yates J., Ballabio A.;
RT "Identification of point mutations in the steroid sulfatase gene of
RT three patients with X-linked ichthyosis.";
RL Am. J. Hum. Genet. 50:483-491(1992).
CC -!- FUNCTION: CONVERSION OF SULFATED STEROID PRECURSORS TO ESTROGENS
CC DURING PREGNANCY.
CC -!- CATALYTIC ACTIVITY: 3-beta-hydroxyandrost-5-en-17-one 3-sulfate +
CC H(2O) = 3-beta-hydroxyandrost-5-en-17-one + sulfate.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SUBCELLULAR LOCATION: MICROSOMAL MEMBRANE. THE SEQUENCE SHOWS
CC SEVERAL MEMBRANE-SPANNING DOMAINS THAT COULD SERVE TO ANCHOR THE
CC PROTEIN IN THE MICROSOMAL MEMBRANE.
CC -!- DISEASE: DEFECTS IN STS ARE A CAUSE OF A VISIBLE PHENOTYPE OF
CC SCALY SKIN, X-LINKED ICHTHYOSIS (XLI).
CC -!- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; J04964; AAA60597.1; -
CC EMBL; M16505; AAA60596.1; -
CC EMBL; M23945; AAA60598.1; -
CC EMBL; M23556; AAA60599.1; -
CC PIR; A32641; A32641.
CC PIR; A25961; A25961.
CC PIR; S05415; S05415.
CC HSP; P15848; IFSU.
CC TIM; 308100; -
CC InterPro; IPR000917; Sulfatase.
CC Pfam; PF00884; Sulfatase; 1.
CC PROSITE; PS00523; SULFATASE_1; 1.
CC PROSITE; PS00149; SULFATASE_2; 1.
CC Hydrolase; Microsome; Transmembrane; Glycoprotein; Steroid metabolism;
CC Pregnancy; Signal; Disease mutation.
CC SIGNAL 1 21 STERYL-SULFATASE
CC CHAIN 22 583 2-AMINO-3-OXOPROPIONIC ACID (BY
CC MOD_RES 75 75 SIMILARITY).
CC ACT_SITE 136 136 POTENTIAL.
CC TRANSMEM 185 206 POTENTIAL.
CC TRANSMEM 213 234 POTENTIAL.
CC CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 459 459 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC VARIANT 341 341 S -> L (IN XLI).
CC /FTID=VAR_007240.
CC VARIANT 372 372 W -> R (IN XLI).
CC /FTID=VAR_007241.
CC C -> Y (IN XLI).
CC /FTID=VAR_007242.
CC FT CONFLICT 23 23 A -> E (IN REF. 2).
CC FT SEQUENCE 583 AA; 65492 MW; 74746AFA9D21A0A6 CRC64;
SQ
Query Match 13.7%; Score 413; DB 1; Length 583;
Best Local Similarity 25.1%; Pred. No. 5.9e-25;
Matches 142; Conservative 78; Mismatches 201; Indels 144; Gaps 22;
QY 67 EPSTTSTSQPLIFLADDDGQFVDVGYGSP- IKTPTLDKLAAGVKL-ENYVQPICTP 124
| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 18 EAESHAASRNILVMDLIGIDGPGCYGNKTIPTNIDRLASGGVKLTQHLAASPLCTP 77
| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 125 SRSQFITGKYIHTGLQH-----SIIRPTQPCNCLPLDNATLPQKLKEVGYSTHMVGKWHL 179
```

```
Db 78 SRAAFMTGRVDRGMAWSRTGVFLFTASSGGGLPTDEITFAKLKQOYSTALIGKWHL 137
Qy 180 GFYRKEC-----MTRRGCFDTFFGSLL-----GSDYYT----- 208
Db 138 GM-----SCHSKTDFCHHPLHHGFNYFYGISLTNLRDCKPGSGSVFTGFKRLVFLPLQIVG 194
Qy 209 -----SCHSKTDFCHHPLHHGFNYFYGISLTNLRDCKPGSGSVFTGFKRLVFLPLQIVG 224
Db 195 VTLLTLAALNCLGLLHVPLGVFFSLLFLAALILFLGLFLHYFRPLNCFM--MRNYELIQ 252
Qy 225 NDNAAWDYNGIYSTOMYTORVOQILASHNPKPIFYIYAOVHSPLOAPGRYFHYRS 284
Db 253 Q---PMSYDN---LTQRLTVEAAQFI-ORNTETPFLVLSVLHVHTALFFSKDFAGKSOH 305
Qy 285 IINTNRRYAAMLSCLEDAINNVTLAKTYGFYNNIIIVSSDNGGQ-----PTAGG 336
Db 306 GV-----YDVAEMDMWSVGQILNLLDELRLANDTLIYFTSDQGAHVEVSSKGEIHGG 359
Qy 337 SNWPLRGSKGYWEGGIRAVGFVHSPLLKNGTVCKEPVHITDWPYTLISLAEGQIDEDI 396
Db 360 SNGIYKGGKANNWEGGIRVPGILRWPRVIOAQKIDETSNWDIFPTVAKLAGAPLPEDR 419
Qy 397 QLDGYDIWETISEGLSRPVD--ILHNIDPIYTRAKNGSWAAGYIWNATQSAIRVOHW 454
Db 420 IIDGRDLMPLE-EG-KSQRSDHEELFYHCNAYLNAVR-----WHPQNSTSI---W 464
Qy 455 KLLTGNPCYSWVPQPSFNSLGNPRWNERITSTGKSV-----WLFNITADPYERVDL 508
Db 465 KAFFFTP-----NFPNPGVSGCFATHVCFGSGSVYTHDPPPLFLFDISKDRERNPL 515
Qy 509 S-----NRYPGIVKLLRRLSOFNKT 529
Db 516 TPASEPREYELKVMQEAADRHQT 540

RESULT 10
ID ARSE_HUMAN STANDARD; PRT; 589 AA.
AC P51690;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Arylsulfatase E precursor (EC 3.1.6.-) (ASE).
GN ARSE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANTS CDPX.
RC TISSUE=Kidney;
RX MEDLINE=95234447; PubMed=7720070;
RA Franco B., Meroni G., Parenti G., Levilliers J., Bernard L.,
RA Gebbia M., Cox L., Maroteaux P., Sheffield L., Rappold G.A.,
RA Andria G., Petit C., Ballabio A.;
RA "A cluster of sulfatase genes on Xp22.3: mutations in
RT chondrodysplasia punctata (CDPX) and implications for warfarin
RT embryopathy.";
RL Cell 81:15-25(1995).
RN [2]
RP VARIANT CDPX TYR-492.
RX MEDLINE=98072521; PubMed=9409863;
RA Parenti G., Buttitta P., Meroni G., Franco B., Bernard L.,
RA Rizzolo M.G., Brunetti-Pierri N., Ballabio A., Andria G.;
RT "X-linked recessive chondrodysplasia punctata due to a new point
RT mutation of the ARSE gene.";
RL Am. J. Med. Genet. 73:139-143(1997).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98163440; PubMed=9497243;
RA Daniele A., Parenti G., D'Addio M., Andria G., Ballabio A.,
RA Meroni G.;
```

```
RT "Biochemical characterization of arylsulfatase E and functional
RT analysis of mutations found in patients with X-linked
RT chondrodysplasia punctata.";
RL Am. J. Hum. Genet. 62:562-577(1998).
CC -1- FUNCTION: MAY BE ESSENTIAL FOR THE CORRECT COMPOSITION OF
CC CARTILAGE AND BONE MATRIX DURING DEVELOPMENT. HAS NO ACTIVITY
CC TOWARD STEROID SULFATES AND SHOWS MAXIMAL ACTIVITY AT NEUTRAL PH.
CC IS ALMOST COMPLETELY INACTIVATED AFTER 10 MINUTES AT 50 DEGREES
CC CELSTUS.
CC -1- ENZYME REGULATION: INHIBITED BY MILLIMOLAR CONCENTRATIONS OF
CC WARFARIN.
CC -1- SUBCELLULAR LOCATION: GOLGI STACK.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE PANCREAS, LIVER, AND KIDNEY.
CC -1- PTM: N-GLYCOSYLATED.
CC -1- DISEASE: DEFECTS IN ARSE ARE THE CAUSE OF X-LINKED RECESSIVE
CC CHONDRODYSPLASIA PUNCTATA (CDPX), A CONGENITAL DEFECT OF BONE AND
CC CARTILAGE DEVELOPMENT CHARACTERIZED BY ABERRANT BONE
CC MINERALIZATION, SEVERE UNDERDEVELOPMENT OF NASAL CARTILAGE, AND
CC DISTAL PHALANGEAL HYPOPLASIA. THIS DISEASE CAN ALSO BE INDUCED BY
CC INHIBITION WITH THE DRUG WARFARIN.
CC -1- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; X83573; CAA58556.1; -
CC HSP; P15848; LFSU.
CC DR
CC MIM; 300180; -
CC DR
CC MIM; 302950; -
CC DR
CC MIM; 302940; -
CC DR
CC InterPro: IPR000917; Sulfatase.
CC Pfam: PF00884; Sulfatase; 1.
CC DR
CC PROSITE; PS00523; SULFATASE_1; 1.
CC DR
CC PROSITE; PS00149; SULFATASE_2; 1.
CC KW Hydrolase; Signal; Glycoprotein; Disease mutation.
CC FT SIGNAL 1 31 POTENTIAL.
CC FT CHAIN 32 589 ARYLSULFATASE E.
CC FT MOD_RES 86 86 POTENTIAL.
CC FT 2-AMINO-3-OXOPROPIONIC ACID (BY
CC SIMILARITY).
CC FT CARBOHYD 58 58 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 125 125 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 258 258 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 344 344 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT VARIANT 12 12 R -> S (IN CDPX).
CC FT /FTID=VAR_007307.
CC FT VARIANT 111 111 R -> P (IN CDPX).
CC FT /FTID=VAR_007308.
CC FT VARIANT 117 117 G -> R (IN CDPX).
CC FT /FTID=VAR_007309.
CC FT VARIANT 137 137 G -> V (IN CDPX).
CC FT /FTID=VAR_007310.
CC FT VARIANT 245 245 G -> R (IN CDPX).
CC FT /FTID=VAR_007311.
CC FT VARIANT 492 492 C -> Y (IN CDPX).
CC FT /FTID=VAR_007312.
CC FT
CC SQ SEQUENCE 589 AA; 65683 MW; 19F7D8DA99E528F CRC64;

Query Match 13.5%; Score 407.5; DB 1; Length 589;
Best Local Similarity 24.0%; Pred. No. 1.6e-24;
Matches 139; Conservative 92; Mismatches 188; Indels 161; Gaps 25;

Qy 66 LEPSTT---STSPHILFIADDDGQFDRVYHGS-IKTPTLDKLAAGVKL-ENVYQP 120
Db 25 LAPSASSDISASRPNILLMADDLGIGDGCYGNVTMTPTNIDRLAEDGVKLTQHTISAAS 84
Qy 121 ICTPSRSQFITGKYQIHTGLQHSI-IRPTQ-----PNCPLDNLATLPQKLKEVGYSTHMGV 175
```

[illegible]

RESULT 11

STBL_ID	STBL_NAME	STANDARD	PRT	577 AA
AC	PI5589;			
DT	01-APR-1990 (Rel. 14, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	Steryl-sulfatase precursor (EC 3.1.6.2) (Steroid sulfatase) (Steryl-sulfate sulfohydrolase) (Arylsulfatase C) (ASC).			
GN	STS.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAT-SPRAGUE-DAWLEY;			
RA	Li X.M., Salido E.C., Gong Y., Yen P.H., Kitada Y., Serikawa T.,			
RA	Shapiro L.J.;			
RL	Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE OF 20-45.			
RC	TISSUE=Liver;			
RX	MEDLINE=89352671; PubMed=2765556;			
RA	Kawano J.-I., Kotani T., Ohtaki S., Minamino N., Matsuo H., Oinuma T.,			
RA	Alkawa E.;			
RT	"Characterization of rat and human steroid sulfatases.";			
RL	Biochim. Biophys. Acta 997:199-205(1989).			
CC	-1- FUNCTION: CONVERSION OF SULFATED STEROID PRECURSORS TO ESTROGENS DURING PREGNANCY.			
CC	-1- CATALYTIC ACTIVITY: 3-beta-hydroxyandrost-5-en-17-one 3-sulfate + H(2)O = 3-beta-hydroxyandrost-5-en-17-one + sulfate.			
CC	-1- SUBUNIT: HOMODIMER.			
CC	-1- SUBCELLULAR LOCATION: MICROSOmal MEMBRANE. THE SEQUENCE SHOWS SEVERAL MEMBRANE-SPANNING DOMAINS THAT COULD SERVE TO ANCHOR THE PROTEIN IN THE MICROSOmal MEMBRANE.			
CC	-1- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.			

	CC	-----	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).		CC	-----
	CC	EMBL; U37138; AAC53097.1; .-			CC	-----
	DR	PIR; S05414; S05414.			DR	
	DR	HSPG; P15848; IFSU.			DR	
	DR	InterPro: IPR000917; Sulfatase.			DR	
	DR	Pfam; PF00884; Sulfatase; 1.			DR	
	DR	PROSITE; PS00523; SULFATASE_1; 1.			DR	
	DR	PROSITE; PS00149; SULFATASE_2; 1.			DR	
	KW	Hydrolase; Microsome; Transmembrane; Glycoprotein; Steroid metabolism;			KW	
	KW	Pregnancy; Signal.			KW	
	FT	SIGNAL	1	19	FT	
	FT	CHAIN	20	577	FT	
	FT	MOD_RES	74	74	FT	
	FT	ACT_SITE	135	135	FT	
	FT	TRANSMEM	182	202	FT	
	FT	TRANSMEM	213	233	FT	
	FT	CARBOHYD	46	46	FT	
	FT	CARBOHYD	332	332	FT	
	FT	CARBOHYD	458	458	FT	
	SQ	SEQUENCE	577 AA;	63678 MW;	SQ	

RESULT 12





```

DR InterPro; IPR000917; Sulfatase.
DR Pfam; PF00884; Sulfatase; 1.
DR PROSITE; PS00523; SULFATASE_1; 1.
DR PROSITE; PS00149; SULFATASE_2; 1.
KW Hydrolase; Complete proteome.
FT ACT_SITE 190 190 POTENTIAL.
FT CONFLICT 191 191 M -> I (IN REF. 2).
SQ SEQUENCE 551 AA; 60717 MW; A04C8BBE5E63E988 CRC64;

Query Match 13.1%; Score 394; DB 1; Length 551;
Best Local Similarity 24.5%; Pred. No. 1.7e-23;
Matches 147; Conservative 97; Mismatches 20; Indels 152; Gaps 29;

QY 2 APRCAGHPPP-----PSPQACVCPGKMLAGALAGFWILCLLTGYLSWGQALREBEA 57
DB 33 ARKGAGYDHPNQYLVPATTIADNMMPVQHPPA-----QDKETQQ 73

QY 58 LLAQAGKLEPSTSTSQPHLIFILADQGRDVGYPHGSEI-----KPTPLDKLAAEGVKL 113
DB 74 KLAELERK-----TGKKPNVVFLLDVGMDVGFNGGAVGNGPTPDIDAVASQGLIL 127

QY 114 ENYVQPTSPRSQFTGKVOIHTGLOHSIIRPT---QPNCLPLDNATLPQKLEKGVYS 170
DB 128 TSAYSQSPSSPTRATILITGQYSH-----HGLMPPMYGQPGGLQ-GUTTLPLLHDQGYV 182

QY 171 THMYGKWLHGFYRKCEMPTRRGDTF--FGSLGSGDYTHYKCDSPGCMGYDLYENDNA 228
DB 183 TOAIGKWHMG-ENKESQPNQVDFDFRGNVS---SDMYTEWR-----DVHVNPEV 229

QY 229 AMDYDNGIYSTQMYTOR-----VOQILASHNP----- 255
DB 230 ALSPDRSEYIKQLPFSKDDVHVRGGEQQAIDITPKYMEDLDQRMWDYGVKFLDKMAKS 289

QY 256 TKPFLFIYAQVHSPLOAQRYPHEHRSI-----ININRRRYAAMLSCLEAINNVTAL 311
DB 290 DKPFLYGTGRGCH-----FDNPNKAYAGSSPARTSYGDCVMEMNDVFANLYKTL 340

QY 312 KTYGFYNNIIYSDDNGQ---PTAGSNWPLRGSGTGWEGGIRAVGFVHSPLLKNKG 368
DB 341 EKNQOLNTLIVFTSDNGPEAEVPHGRT--PFRGAGSTWEGGVVPTFYVW-----KG 393

QY 369 TVCKEP-----VHTDWPYPLISLARGIDEDTQLDGYDIWETISGLRSRPRVDILHNI 422
DB 394 MI--QPRKSDIGVDLADLFTALDLA-GH-----PGAQVANLV-----PKTTFIDGV 437

QY 423 DP----IYTRAKNGSWAGYGIWNTATQSAIRVQHWKLLTGNPCYSQVWV-PPQSFSLGP 477
DB 438 DQTSFFLTNGQSQRKAHEHYFL--NGKLAAVRMDEFK-----YHVLIOQPYAYTOSGY 488

QY 478 NRWNERITSTGKSVWLFNITADPYRVDLSNRYPGIVKLLRLSQFNKTAVPVRYPP 537
DB 489 QGGFTGIVMOTAGSSV--FNLYTPQESDSIGVRHHPGVLPQTEMHAYME--ILKYYPP 544

QY 538 K 538
DB 545 R 545

RESULT 14
STS_MOUSE STANDARD; PRT; 624 AA.
AC P50427;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Steryl-sulfatase precursor (EC 3.1.6.2) (Steroid sulfatase) (Steryl-
DE sulfatase sulfohydrolase) (Arylsulfatase C) (ASC).
GN STS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;

```

```

Db 302 FADPGFAGRSL-----HGAYGDSVEEMDVGVRVLAALDELGLARETLVYFTSDHGAHVEE 357
Qy 332 -----PTAGGSGNWPLRGSGKGYWEGGIRAVGFVHSPILKKNKGTVCKEPVHTTDMVPTLIS 386
Db 358 LQPRGERGGSGVFRGGKGNWEGGVPCVLRWPRELSGCRVAEPTSLMDVFPVVAR 417
Qy 387 LAEQIDIEDIQDGYDIWETISEGLRSPRVDILHNIDPIYTKAKNGSWAAGYGIWNTAIQ 446
Db 418 LAGAELPDGRVIDRDLMLPLLRGDAQRSEHEFLFYHCNAYLQAVR--WHNG----- 466
Qy 447 SAIRVOHKLITGNPGYSDWVPSFSGMLGNRWHNERITSTGKS-----VWLEFNIT 499
Db 467 SAV-----WKAFYFPTP-----NFAPAGANGCFSTHVCILCAGPAHVTAHPDPLLFDLT 513
Qy 500 ADPYERVDLS 509
Db 514 RDPGERRPLT 523

RESULT 15
ARSA_HUMAN
ID ARSA_HUMAN STANDARD; PRT; 507 AA.
AC P15289;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Arylsulfatase A precursor (EC 3.1.6.8) (ASA) (Cerebroside-sulfatase).
GN ARSA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90361046; PubMed=1975241;
RA Kreyling J., von Figura K., Gieselmann V.;
RT "Structure of the arylsulfatase A gene.";
RL Eur. J. Biochem. 191:627-631(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89093115; PubMed=2562955;
RA Stein C., Gieselmann V., Kreyling J., Schmidt B., Pohlmann R.,
RA Waheed A., Meyer H.E., O'Brien J.S., von Figura K.;
RT "Cloning and expression of human arylsulfatase A.";
RL J. Biol. Chem. 264:1252-1259(1989).
RN [3]
RP SEQUENCE FROM N.A.
RA Adams M.D., Kerlavage A.R., Fuldner R.A., Phillips C.A., Venter J.C.;
RT "Complete sequence of a chromosome 22q subtelomeric BAC.";
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 19-33 AND 434-479, AND SUBUNITS.
RX MEDLINE=92338230; PubMed=1352993;
RA Fujii T., Kobayashi T., Honke K., Gasa S., Ishikawa M., Shimizu T.,
RA Makita A.;
RT "Proteolytic processing of human lysosomal arylsulfatase A.";
RL Biochim. Biophys. Acta 1122:93-98(1992).
RN [5]
RP PARTIAL SEQUENCE, AND 2-AMINO-3-OXOPROPIONIC ACID MODIFICATION SITE.
RX MEDLINE=95354208; PubMed=7628016;
RA Schmidt B., Selmer T., Ingendoh A., von Figura K.;
RT "A novel amino acid modification in sulfatases that is defective in
multiple sulfatase deficiency.";
RL Cell 82:271-278(1995).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
RX MEDLINE=98191325; PubMed=9521684;
RA Lukatela G., Krauss N., Theis K., Selmer T., Gieselmann V.,
RA von Figura K., Saenger W.;
RT "Crystal structure of human arylsulfatase A: the aldehyde function
and the metal ion at the active site suggest a novel mechanism for
sulfate ester hydrolysis.";

```

```

RL Biochemistry 37:3654-3664(1998).
RN [7]
RP REVIEW ON MLD VARIANTS.
RX MEDLINE=95170731; PubMed=7866401;
RA Gieselmann V., Zlotogora J., Harris A., Wenger D.A., Morris C.P.;
RT "Molecular genetics of metachromatic leukodystrophy.";
RL Hum. Mutat. 4:233-242(1994).
RN [8]
RP VARIANT MLD GLN-84.
RX MEDLINE=92344341; PubMed=1353340;
RA Kappler J., von Figura K., Gieselmann V.;
RT "Late-onset metachromatic leukodystrophy: molecular pathology in two
siblings.";
RL Ann. Neurol. 31:256-261(1992).
RN [9]
RP VARIANT MLD PHE-96.
RX MEDLINE=91328147; PubMed=1678251;
RA Gieselmann V., Fluhrty A.L., Tonnesen T., von Figura K.;
RT "Mutations in the arylsulfatase A pseudodeficiency allele causing
metachromatic leukodystrophy.";
RL Am. J. Hum. Genet. 49:407-413(1991).
RN [10]
RP VARIANT MLD ASP-99.
RX MEDLINE=91206410; PubMed=1673291;
RA Kondo R., Wakamatsu N., Yoshino H., Fukuhara N., Miyatake T.,
RA Tsuji S.;
RT "Identification of a mutation in the arylsulfatase A gene of a
patient with adult-type metachromatic leukodystrophy.";
RL Am. J. Hum. Genet. 48:971-978(1991).
RN [11]
RP VARIANT MLD LEU-426, AND VARIANTS CYS-193 AND SER-391.
RX MEDLINE=91074201; PubMed=1670590;
RA Polten A., Fluhrty A.L., Fluhrty C.B., Kappler J., von Figura K.,
RA Gieselmann V.;
RT "Molecular basis of different forms of metachromatic leukodystrophy.";
RL New Engl. J. Med. 324:18-22(1991).
RN [12]
RP VARIANT MLD SER-122.
RX MEDLINE=94063853; PubMed=7902317;
RA Honke K., Kobayashi T., Fujii T., Gasa S., Xu M., Takamaru Y.,
RA Kondo R., Tsuji S., Makita A.;
RT "An adult-type metachromatic leukodystrophy caused by substitution of
serine for glycine-122 in arylsulfatase A.";
RL Hum. Genet. 92:451-456(1993).
RN [13]
RP VARIANT MLD LEU-136.
RX MEDLINE=95163939; PubMed=7860068;
RA Kafert S., Heinisch U., Zlotogora J., Gieselmann V.;
RT "A missense mutation p136L in the arylsulfatase A gene causes
instability and loss of activity of the mutant enzyme.";
RL Hum. Genet. 95:201-204(1995).
RN [14]
RP VARIANT MLD ARG-245.
RX MEDLINE=93319632; PubMed=8101083;
RA Hasegawa Y., Kawame H., Eto Y.;
RT "Mutations in the arylsulfatase A gene of Japanese patients with
metachromatic leukodystrophy.";
RL DNA Cell Biol. 12:493-498(1993).
RN [15]
RP VARIANT MLD MET-274.
RX MEDLINE=94004907; PubMed=8104633;
RA Harvey J.S., Nelson P.V., Carey W.F., Robertson E.F., Morris C.P.;
RT "An arylsulfatase A (ARSA) missense mutation (T274M) causing late-
infantile metachromatic leukodystrophy.";
RL Hum. Mutat. 2:261-267(1993).
RN [16]
RP VARIANT MLD SER-309.
RX MEDLINE=93318834; PubMed=8101038;
RA Kreyling J., Böhne W., Bosenberg C., Marchesini S., Turpin J.C.,
RA Baumann N., von Figura K., Gieselmann V.;
RT "High residual arylsulfatase A (ARSA) activity in a patient with
late-infantile metachromatic leukodystrophy.";
RL Am. J. Hum. Genet. 53:339-346(1993).

```

RN [17]  
 RP VARIANT SER-350.  
 RX MEDLINE=90083282; PubMed=2574462;  
 RA Gieselmann V., Polten A., Kreysing J., von Figura K.;  
 RT "Arylsulfatase A pseudodeficiency: loss of a polyadenylation signal  
 and N-glycosylation site."; *Hum. Genet.* 91:73-77(1993).  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:9436-9440(1989).  
 RN [18]  
 RP VARIANT MLD LEU-426.  
 RX MEDLINE=93202658; PubMed=8095918;  
 RA Barth M.L., Fensom A., Harris A.;  
 RT "Prevalence of common mutations in the arylsulphatase A gene in  
 metachromatic leukodystrophy patients diagnosed in Britain."; *Hum. Genet.* 91:73-77(1993).  
 RL Hum. Genet. 91:73-77(1993).  
 RN [19]  
 RP VARIANTS MLD VAL-212; VAL-224 AND TYR-295.  
 RX MEDLINE=94154687; PubMed=7906588;  
 RA Barth M.L., Fensom A., Harris A.;  
 RT "Missense mutations in the arylsulphatase A genes of metachromatic  
 leukodystrophy patients."; *Hum. Mol. Genet.* 2:2117-2121(1993).  
 RL Hum. Mol. Genet. 2:2117-2121(1993).  
 RN [20]  
 RP VARIANTS MLD LEU-82; TYR-172; CYS-201; GLN-311; VAL-335 AND TRP-390.  
 RX MEDLINE=96047150; PubMed=7581401;  
 RA Barth M.L., Fensom A., Harris A.;  
 RT "Identification of seven novel mutations associated with  
 metachromatic leukodystrophy."; *Hum. Mutat.* 6:170-176(1995).  
 RL Hum. Mutat. 6:170-176(1995).  
 RN [21]  
 RP CHARACTERIZATION OF VARIANTS MET-274 AND VAL-335.  
 RX MEDLINE=96303701; PubMed=8723680;  
 RA Hess B., Kafert S., Heinisch U., Wenger D.A., Zlotogora J.,  
 Gieselmann V.;  
 RT "Characterization of two arylsulphatase A missense mutations D335V and  
 T274M causing late infantile metachromatic leukodystrophy."; *Hum. Mutat.* 7:311-317(1996).  
 RL Hum. Mutat. 7:311-317(1996).  
 RN [22]  
 RP VARIANTS MLD.  
 RX MEDLINE=97245886; PubMed=9090526;  
 RA Draghia R., Letourneur F., Dragan C., Manicom J., Blanchot C.,  
 Kahn A., Poenaru L., Caillaud C.;  
 RT "Metachromatic leukodystrophy: identification of the first deletion in  
 exon 1 and of nine novel point mutations in the arylsulphatase A  
 gene."; *Hum. Mutat.* 9:234-242(1997).  
 RL Hum. Mutat. 9:234-242(1997).  
 RN [23]  
 RP VARIANT MLD 406-SER--THR-408 DEL.  
 RX MEDLINE=98141126; PubMed=9490297;  
 RA Regis S., Flocamo M., Stroppiano M., Corsolini F., Caroli F.,  
 Gatti R.;  
 RT "A 9-bp deletion (2320del19) on the background of the arylsulphatase A  
 pseudodeficiency allele in a metachromatic leukodystrophy patient and  
 in a patient with nonprogressive neurological symptoms."; *Hum. Genet.* 102:50-53(1998).  
 RL Hum. Genet. 102:50-53(1998).  
 RN [24]  
 RP VARIANTS MLD PRO-135 AND SER-179.  
 RX MEDLINE=98260871; PubMed=9600244;  
 RA Gomez-Lira M., Perusi C., Mottes M., Pignatti P.F., Manfredi M.,  
 Rizzuto N., Salvati A.;  
 RT "Molecular genetic characterization of two metachromatic  
 leukodystrophy patients who carry the T799G mutation and show  
 different phenotypes; description of a novel null-type mutation."; *Hum. Genet.* 102:459-463(1998).  
 RL Hum. Genet. 102:459-463(1998).  
 RN [25]  
 RP VARIANTS MLD GLN-390 AND TYR-397.  
 RX MEDLINE=98112481; PubMed=9452102;  
 RA Coulter-Mackie M.B., Gagnier L.;  
 RT "Two novel mutations in the arylsulphatase A gene associated with  
 juvenile (R390Q) and adult onset (H397Y) metachromatic  
 leukodystrophy."; *Hum. Mutat. Suppl.* 1:S254-S256(1998).  
 RL Hum. Mutat. Suppl. 1:S254-S256(1998).  
 RN [26]  
 RP VARIANT HIS-496.

RX MEDLINE=98415722; PubMed=9744473;  
 RA Ricketts M.H., Poretz R.D., Manowitz P.;  
 RT "The R496H mutation of arylsulphatase A does not cause metachromatic  
 Query Match 12.5%; Score 378; DB 1; Length 507;  
 Best Local Similarity 26.1%; Pred. No. 2.8e-22;  
 Matches 128; Conservative 75; Mismatches 195; Indels 92; Gaps 19;  
 QY 76 PHLIFILADDOGFRDVGVG-SEIKTPTLDKLAAGVKLENYVO-PICTPSRSQFTGK 133  
 Db 21 PNIWLIIFADDLGYGDLGCIHPSSSTPNLDQLAAGURFTDFYVPVSLCTPSRAALLTGR 80  
 QY 134 YQIHTGLQHSIIRPTQPNCLDNLATLPQKLKEVGYSTHVMGKWHLGFYRKEC-MPTRRG 192  
 Db 81 LPVRMGMYPGVLVPSSRGGGLPEEVTVAEVLAAARGYLTMAGKWHLGVGPEGAFLPHQG 140  
 QY 193 FDTFFGSLGSGDYIYTHYK-----CDSPQCMCGYDLYENDNAAWDYDNGIYS 238  
 Db 141 FHRELGII-----PYSHDQGPCQLTCFPPATPCD--GGC-----DQGLVP 178  
 QY 239 TQM-----YTORVQOILA-SHNPTKPIFLVIAYOAVHSPLOAQGRY 278  
 Db 179 IPLLANLSVEAQPPWLPGLLEARYMAFAHDLMAADARQDRPFELIYASHHTYP-QFSGQS 237  
 QY 279 FEHYRSIINRRRYAAMLSCLEAINNVTIALKTYGYFNNSIIIIYSSDNGQP---TAG 335  
 Db 238 FAE-RS---GRGPFGLDLMELDAAVGTLMTAIGDLGLLEETLVIFTADNGPETMRMSRG 292  
 QY 336 GSNWPLRSGKTYWEGGIRAVGFVHSPLLKNKGVCKVEPVHITDWYPTLISLAEQIDED 395  
 Db 293 GCSGLLRGCGKGTYYEGGVREPALAFWFCHTIAPG-VTHELASSLDLPTLAALAGAPL-PN 350  
 QY 396 IOLGGRDIWETISBGLRSPRVDILHNIDPIYTKAKNGSWAAGYIGWNTAIQSAIRVQHWK 455  
 Db 351 VTLDGFDLSPLLLGTGKSPQSLF--FYPSYDPDEVRGVFAVRTGKYKA-----H 397  
 QY 456 LLTGNPGYSDWVPPQSFNSLGNPNRWHNERITSTGKSVWLENITADPYERVDLSNRYPGI 515  
 Db 398 FFTQGSASHD-----TTADPACHASSSLTAH--EPELLYDLKSDPCGENTLLGGVAGA 448  
 QY 516 VKLLLRRLSQ 525  
 Db 449 TPEVLQALKQ 458

Search completed: July 23, 2002, 19:07:50  
 Job time: 366 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 15:05:29 ; Search time 356.89 Seconds  
(without alignments)  
10901.197 Million cell updates/sec

Title: US-09-495-823-8

Perfect score: 2266

Sequence: 1 caccgctcgccacgcgtc.....tgcacctggtgcgaattc 2266

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- N\_Geneseq\_032802.\*
- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
  - 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
  - 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
  - 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
  - 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
  - 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
  - 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
  - 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
  - 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
  - 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
  - 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
  - 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
  - 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
  - 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
  - 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
  - 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
  - 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
  - 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
  - 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
  - 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
  - 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
  - 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
  - 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
  - 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2234.2	98.6	2253	22	AAH46863 Human 23553 sulfat
2	2033.8	89.8	4639	22	AAS45962 Human DNA encoding
3	2033.8	89.8	4640	21	AAC78483 Human PRO708 (UNO3
4	2033.8	89.8	4650	20	AAZ33987 Human PRO708 nucle
5	1781.6	78.6	1800	22	AAH46867 Human 23553 sulfat
6	890	39.3	1375	22	AAS22726 Human cDNA encodin
7	736.2	32.5	1340	22	AAS22490 Human cDNA encodin
8	625.2	27.6	729	21	AAK02344 Human colon cancer
9	596.2	26.3	2940	22	AAH46862 Human 25278 sulfat

10	594.4	26.2	1710	22	AAH46866 Human 25278 sulfat
11	575.2	25.4	629	22	AAS41205 cDNA encoding nove
12	538.4	-23.8	562	22	AAH99356 Human protein enco
13	210.4	9.3	1615	22	AAS28640 Genomic sequence #
14	139.6	6.2	1082	21	AAK02370 Human colon cancer
15	129	5.7	1740	23	ABL12859 Drosophila melanog
16	123.2	5.4	294	21	AAA00278 Human colon cancer
17	110.6	4.9	1629	23	ABL14291 Drosophila melanog
18	98.2	4.3	249	19	AAK11642 Human biallelic po
19	98.2	4.3	249	19	AAK12813 Human biallelic po
20	88.2	3.9	3039	23	ABL09675 Drosophila melanog
21	55.6	2.5	4030	23	ABL12884 Drosophila melanog
22	55.4	2.4	4036	23	ABL14290 Drosophila melanog
23	53	2.3	10732	21	AAA10594 Gene encoding a su
24	52.4	2.3	2401	20	AAZ20800 Human steroid sulp
25	52	2.3	53	20	AAC23390 Human PRO708 hybr
26	52	2.3	53	21	AAC78675 Human PRO708 hybr
27	52	2.3	11473	24	ABL33354 Human immune syste
28	51.2	2.3	1461	23	ABL12885 Drosophila melanog
29	50.6	2.2	4006	23	ABL12858 Drosophila melanog
30	48.2	2.1	5856	23	ABL09674 Human immune syste
31	47.4	2.1	6090	24	ABL33578 Human immune syste
32	47	2.1	249	18	AAT89608 Hepatitis C virus
33	46.4	2.0	876	21	AAC59201 Human secreted pro
34	45.2	2.0	431	22	AAI87912 Human polynucleoti
35	45.2	2.0	3973	24	ABL32171 Human immune syste
36	45.2	2.0	16167	24	ABL33082 Human immune syste
37	45.2	2.0	16167	24	ABL34528 Human metastasis a
38	44.8	2.0	382	22	AAT80370 Human polynucleoti
39	44.8	2.0	6436	24	ABL32681 Human immune syste
40	44.6	2.0	583	22	AAS29116 cDNA encoding for
41	44.6	2.0	700	22	AAH92784 Human inflammatory
42	44.6	2.0	700	22	AAH92785 Human inflammatory
43	44.6	2.0	4237	19	AAV61487 Human secreted pro
44	44.4	2.0	882	21	AAZ52528 Human secreted pro
45	44.4	2.0	997	22	AAC91493 Human PRO6006 cDNA

ALIGNMENTS

RESULT 1

AAH46863

ID AAH46863 standard; cDNA; 2253 BP.

XX

XX AAH46863;

XX AC

XX 25-SEP-2001 (first entry)

DT

XX Human 23553 sulfatase polypeptide encoding cDNA.

DE

XX Sulfatase; 23553 sulfatase; nontropic; neuroprotective; antibacterial;

KW antiinflammatory; vasotropic; antitumor; gene therapy; human; ss.

KW

XX Homo sapiens.

OS

XX Key Location/Qualifiers

FH CDS 324..2123

FT /\*tag= a

FT /product= "23553 sulfatase"

XX

PN WO200155411-A2.

XX

PD 02-AUG-2001.

XX

PF 31-JAN-2001; 2001WO-US03266.

XX

PR 31-JAN-2000; 2000US-0495823.

XX

PA (MILL-) MILLENNIUM PHARM INC.

XX

PI Glucksmann MA, Williamson M, Rudolph-Owen LA, Tsai F;

XX







Db	649	ggaggaggaggagaaagtgaatgtgtctggagaaagagcgagccctcttctgtttcttcgga	708
Qy	142	gtcccatcattaaagccatcaactctctggaagattaaaagtgtcggacaatggtgacagctg	201
Db	709	gtcccatcattaaagccatcaactctctggaagattaaaagtgtcggacaatggtgacagctg	768
Qy	202	agaggagaggagattctctgccaggtggagagtcttcaccgtctctgttgggtgcaatgtgt	261
Db	769	agaggagaggagattctctgccaggtggagagtcttcaccgtctgttgggtgcaatgtgt	828
Qy	262	gcaccgcacagcgcgggcggtgttctccgctggagattcaactgggacacctgagt	321
Db	829	gcgccgcagcgcgcgggcggtgttctccgctggagattcaactgggacacctgagt	888
Qy	322	gaatggctccacagggctgtgcggggcatccgcctccgcctctccacaggcctgtgtct	381
Db	889	gaatggctccacagggctgtgcggggcatccgcctccgcctctccacaggcctgtgtct	948
Qy	382	gtcctggaaagatgctagcaatggggggcgctgaggattctgtgatcctctgctctctca	441
Db	949	gtcctggaaagatgctagcaatggggggcgctgaggattctgtgatcctctgctctctca	1008
Qy	442	cttatggttacctgtccttggggcaggccttagaagaagagaagaaggggccttactag	501
Db	1009	cttatggttacctgtccttggggcaggccttagaagaagagaagaaggggccttactag	1068
Qy	502	ctcaagctggagagaaactagagccagcacaaactccacctccagcccccactctcaatt	561
Db	1069	ctcaagctggagagaaactagagccagcacaaactccacctccagcccccactctcaatt	1128
Qy	562	tcactctagcggatgatacaggagatttagagatgtgggttaccccggtctgagattaaaa	621
Db	1129	tcactctagcggatgatacaggagatttagagatgtgggttaccccggtctgagattaaaa	1188
Qy	622	caactctcttgacaagctcgctgcggaaaggagttaaactggagaactactatgtccagc	681
Db	1189	caactctcttgacaagctcgctgcggaaaggagttaaactggagaactactatgtccagc	1248
Qy	682	ctatttgcacaccatccagaggtcagtttattactggaaagtatcagatatacacacggac	741
Db	1249	ctatttgcacaccatccagaggtcagtttattactggaaagtatcagatatacacacggac	1308
Qy	742	ttcaacattctataaagacctaaccacccaactgtttacctcttgacaaatggcaaccc	801
Db	1309	ttcaacattctataaagacctaaccacccaactgtttacctcttgacaaatggcaaccc	1368
Qy	802	tacctgaagaactggaaggaggttgatattcaacgcataatggtcggaaaaatggcaattgg	861
Db	1369	tacctgaagaactggaaggaggttgatattcaacgcataatggtcggaaaaatggcaattgg	1428
Qy	862	gtttttacagaaaagaatgcatgcccacagagaaggatttgatacctcttttttggttccc	921
Db	1429	gtttttacagaaaagaatgcatgcccacagagaaggatttgatacctcttttttggttccc	1488
Qy	922	tttttgggaagtggggattactatacacatacaaatgtgacagtcctctgggatgtgtggct	981
Db	1489	tttttgggaagtggggattactatacacatacaaatgtgacagtcctctgggatgtgtggct	1548
Qy	982	atgacttgtataaaaaacgacaatgctgctgggactatgacaaatgggataatactccaac	1041
Db	1549	atgacttgtataaaaaacgacaatgctgctgggactatgacaaatgggataatactccaac	1608
Qy	1042	agatgtacactcagagagtacagcaaatcttagcttcccataaaccccaaaagcctatat	1101
Db	1609	agatgtacactcagagagtacagcaaatcttagcttcccataaaccccaaaagcctatat	1668
Qy	1102	tttttatatatgtcctatcaagctgttccaattccacctgcaagctcctctggcaggtatttcg	1161
Db	1669	tttttatatatgtcctatcaagctgttccaattccacctgcaagctcctctggcaggtatttcg	1728
Qy	1162	aaacactacogatccattatacaacataaaacagaggagagatagtgtgcattgtcttctctgt	1221
Db	1729	aaacactacogatccattatacaacataaaacagaggagagatagtgtgcattgtcttctctgt	1788

Qy	1222	tagatgaagcaatacaacaa	cgtgacatctggctctaaagacttatggtttctatacaaca	1281
Db	1789	tagatgaagcaatacaacaa	cgtgacatctggctctaaagacttatggtttctatacaaca	1848
Qy	1282	gcattatcatttactctcaga	taattgggtggccagcctacggcagagggagtaactggcc	1341
Db	1849	gcattatcatttactctcaga	taattgggtggccagcctacggcagagggagtaactggcc	1908
Qy	1342	ctctcagaggtagcaaaaggaa	catacttggaagaggagatccgggctctaggcttttggc	1401
Db	1909	ctctcagaggtagcaaaaggaa	catacttggaagaggagatccgggctctaggcttttggc	1968
Qy	1402	atagcccaactcttgaaaaa	caagggaacagtggttaaggaaacctgtgcacatcaactgact	1461
Db	1969	atagcccaactcttgaaaaa	caagggaacagtggttaaggaaacctgtgcacatcaactgact	2028
Qy	1462	ggtaccccaactctcattt	ccactggctggaagacagatgtatgaggacattcaactagatg	1521
Db	2029	ggtaccccaactctcattt	ccactggctggaagacagatgtatgaggacattcaactagatg	2088
Qy	1522	gctatgatatctggagacata	aaagttaggggttctgctccaccgcagtagatatatttgc	1581
Db	2089	gctatgatatctggagacata	aaagttaggggttctgctccaccgcagtagatatatttgc	2148
Qy	1582	ataacattgaccccatata	caccaaggcaaaaaatggctctctggcagcagagctatggga	1641
Db	2149	ataacattgacccc-tata	caccaaggcaaaaaatggctctctggcagcagagctatggga	2207
Qy	1642	tctggaacactgcaatccag	tcagccatcagagtgccagcactggaaattgtcttacaggaa	1701
Db	2208	tctggaacactgcaatccag	tcagccatcagagtgccagcactggaaattgtcttacaggaa	2267
Qy	1702	atcttggtcagcgactgggt	ccccccctcagcttttccagaaacctgggacccagccggt	1761
Db	2268	atcttggtcagcgactgggt	ccccccctcagcttttccagaaacctgggacccagccggt	2327
Qy	1762	ggcacaatgaacggatcac	ctctgcactggcaaaagtgtatggcttttcaacatccacag	1821
Db	2328	ggcacaatgaacggatcac	ctctgcactggcaaaagtgtatggcttttcaacatccacag	2387
Qy	1822	ccgaccataatgagaggg	tggaaacctatctaaacaggtatccagaagaactgtgaagaagctcc	1881
Db	2388	ccgaccataatgagaggg	tggaaacctatctaaacaggtatccagaagaactgtgaagaagctcc	2447
Qy	1882	tacggaggctctcacgtt	caacaaactgcagtgccggtcaggtatccccccaaagacc	1941
Db	2448	tacggaggctctcacgtt	caacaaactgcagtgccggtcaggtatccccccaaagacc	2507
Qy	1942	ccagaagttaaccttaggct	caattggagggggtctagggaacctatggtatagagaggaaacca	2001
Db	2508	ccagaagttaaccttaggct	caattggagggggtctagggaacctatggtatagagaggaaacca	2567
Qy	2002	agaaaaagaaagccaaag	caaaaaatcaggcttgagaaaaagcaaaagaaaaagc-aaaaaaga	2060
Db	2568	agaaaaagaaagccaaag	caaaaaatcaggcttgagaaaaagcaaaagaaaaaaga	2627
Qy	2061	agaagaaacagcagaagc	agctctcaggttccaacttcaggttgcattcaggt 2107	
Db	2628	agaagaaacagcagaagc	agctctcaggttccaacttcaggttgcattcaggt 2674	

### RESULTS

·RESUL  
AAC78

AAC78483  
ID AAC78483 standard: cDNA: 4640 BP.

AAC78483;

AC  
XX  
AAC/04037

DT 08-FEB-2001 (first entry)

XX  
DE Human PRO708 (UNQ372) nucleotide sequence SEQ ID NO:113.

Human; secreted protein; transmembrane protein; PRO; EST; cytosolic;



Db	1789	tagatgaagcaatcaacaacgtgacatggctctctaaagacttatgggtcttcatacaaca	1848
Qy	1282	gcattatcatttactcttcagataatgggtggccagccttacgcgagggaggaactcggc	1341
Db	1849	gcattatcatttactcttcagataatgggtggccagccttacgcgagggaggaactcggc	1908
Qy	1342	ctctcagaggtgcaaaagaaacatatgtgggaaggaggatccgggctgtaggcctttgtgc	1401
Db	1909	ctctcagaggtgcaaaagaaacatatgtgggaaggaggatccgggctgtaggcctttgtgc	1968
Qy	1402	atagcccaactctgaaaaacaaggaaacagtggtgaaggaaacctgtgcacatcaactgact	1461
Db	1969	atagcccaactctgaaaaacaaggaaacagtggtgaaggaaacctgtgcacatcaactgact	2028
Qy	1462	ggtaaccccaactctcatttcaactggctggaaggacagatgatgaggacaattcaactagatg	1521
Db	2029	ggtaaccccaactctcatttcaactggctggaaggacagatgatgaggacaattcaactagatg	2088
Qy	1522	gctatgatattctggagaaacataaagtgggggtctctgcctaccoccgagtagatatattgc	1581
Db	2089	gctatgatattctggagaaacataaagtgggggtctctgcctaccoccgagtagatatattgc	2148
Qy	1582	atacaattgaaccccatatacaccaaggcaaaaaatggctctggcagcaggctataggga	1641
Db	2149	atacaattgaacccc-tatacaccaaggcaaaaaatggctctggcagcaggctataggga	2207
Qy	1642	tctggaaacactgcgaatccagctcagccatcagagtgacagctggaaatgtcttacaggaa	1701
Db	2208	tctggaaacactgcgaatccagctcagccatcagagtgacagctggaaatgtcttacaggaa	2267
Qy	1702	atcctggctacagcgactgggtctcccccctcagctcttcagcaacctgggaccgaaccggt	1761
Db	2268	atcctggctacagcgactgggtctcccccctcagctcttcagcaacctgggaccgaaccggt	2327
Qy	1762	ggcacaatgaacggatcaactcgtcaactggcaaaagtgtatggcttttcaacatcacag	1821
Db	2328	ggcacaatgaacggatcaactcgtcaactggcaaaagtgtatggcttttcaacatcacag	2387
Qy	1822	ccgaccacatgatgagagggtgggacctatcttaacaggtatccagggaatcgtgaaagaagctcc	1881
Db	2388	ccgaccacatgatgagagggtgggacctatcttaacaggtatccagggaatcgtgaaagaagctcc	2447
Qy	1882	tacgggggtctctcaacagttccaacaaatcgatgcgggtcagtgatcccccccaagacc	1941
Db	2448	tacgggggtctctcaacagttccaacaaatcgatgcgggtcagtgatcccccccaagacc	2507
Qy	1942	ccagaagtaaaccttaggctcaatgagggggtctagggaacctggtatagagaggaaacca	2001
Db	2508	ccagaagtaaaccttaggctcaatgagggggtctagggaacctggtatagagaggaaacca	2567
Qy	2002	agaaaaagaagccaagcaaaaatcagggttgaaaaagcaaaaagcaaaaagaaagc-aaaaaaga	2060
Db	2568	agaaaaagaagccaagcaaaaatcagggttgaaaaagcaaaaagcaaaaagcaaaaagaa	2627
Qy	2061	agaagaaacagcagaaagcagtcctcaggttcaacttgccattcagggt 2107	
Db	2628	agagaaacagcagaaagcagtcctcaggttcaacttgccattcagggt 2674	

RESULT	4
AAZ33987	
ID	AAZ33987 standard; CDNA; 4650 BP.
XX	
AC	AAZ33987;
XX	
DT	07-DEC-1999 (first entry)
XX	
DE	Human PRO708 nucleotide sequence.
XX	
KW	Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW	probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW	secreted protein; transmembrane protein; ss.



Db 1909 cctcagaggtagcaaggaacataattggaaggaggatccggctgtaggctttgtgc 1968  
Qy 1402 atagccacttctgaaacaaaggaaacagtggtgtaaggaaacctgtcacatcactgact 1461  
Db 1969 atagccacttctgaaacaaaggaaacagtggtgtaaggaaacctgtcacatcactgact 2028  
Qy 1462 ggtaccacactctcatttctcactggtgctgaaggacagattgatgagacattcaactagatg 1521  
Db 2029 ggtaccacactctcatttctcactggtgctgaaggacagattgatgagacattcaactagatg 2088  
Qy 1522 gctatgatattctggagaccataaagtgggtcttcctgcctcaaccccgagtagatatttgc 1581  
Db 2089 gctatgatattctggagaccataaagtgggtcttcctgcctcaaccccgagtagatatttgc 2148  
Qy 1582 ataacattgacccctatacaccaaggcaaaaaatggctctggtcgcagcaggctatgga 1641  
Db 2149 ataacattgacccc-tatacaccaaggcaaaaaatggctctggtcgcagcaggctatgga 2207  
Qy 1642 tctgaaacactgcaatccagtcagccatcagagtcagcactggaattgcttacaggaa 1701  
Db 2208 tctgaaacactgcaatccagtcagccatcagagtcagcactggaattgcttacaggaa 2267  
Qy 1702 atcctggctacagcactgggttccccctcagcttcttcagcaacctgggacccgaggt 1761  
Db 2268 atcctggctacagcactgggttccccctcagcttcttcagcaacctgggacccgaggt 2327  
Qy 1762 ggcacaaatgaacggatcacctcgtcaactgcaactgcaactggaattgctttcaacatcacag 1821  
Db 2328 ggcacaaatgaacggatcacctcgtcaactgcaactggaattgctttcaacatcacag 2387  
Qy 1822 cgcacccatagaggggtggacctatctaacaggtatccaggaatcgtgaagaagctcc 1881  
Db 2388 cgcacccatagaggggtggacctatctaacaggtatccaggaatcgtgaagaagctcc 2447  
Qy 1882 tacgagagctctacagttcaacaaactgacgtgcggtcaggtatccccccaaagacc 1941  
Db 2448 tacgagagctctacagttcaacaaactgacgtgcggtcaggtatccccccaaagacc 2507  
Qy 1942 ccagagtaaaccttagctcaatgaggggtcttagggaccatggtatagagagaaacca 2001  
Db 2508 ccagagtaaaccttagctcaatgaggggtcttagggaccatggtatagagagaaacca 2567  
Qy 2002 agaaaaaagaagcaagcaaaatcaggctgagaaaaagcaaaagaaaaagc-aaaaaaaga 2060  
Db 2568 agaaaaaagaagcaagcaaaatcaggctgagaaaaagcaaaagaaaaaaaga 2627  
Qy 2061 agaaaaaagcagcaaaagcagctcaggttcaacttgcattcaggt 2107  
Db 2628 agaaaaaagcagcaaaagcagctcaggttaaacccagcaaatgtggt 2674

RESULT 5  
AAH46867  
ID AAH46867 standard; cDNA; 1800 BP.  
AC  
XX  
AC AAH46867;  
XX

DT 25-SEP-2001 (first entry)  
XX  
DE Human 23553 sulfatase polypeptide coding sequence.  
XX

KW Sulfatase; 23553 sulfatase; nootropic; neuroprotective; antibacterial;  
KW antiinflammatory; vasotropic; antitumor; gene therapy; human; ss.  
XX

OS Homo sapiens.

XX Key Location/Qualifiers  
FH 1..1800  
CDS /\*tag= a  
FT /product= "23553 sulfatase"  
XX

PN W0200155411-A2.

XX 02-AUG-2001.  
PD  
XX  
XX 31-JAN-2001; 2001WO-US03266.  
XX  
XX 31-JAN-2000; 2000US-0495823.  
XX  
XX (MILL-) MILLENNIUM PHARM INC.  
PI Glucksmann MA, Williamson M, Rudolph-Owen LA, Tsai F;  
XX  
XX WPI; 2001-476214/51.  
DR P-PSDB; AAB85483.  
XX  
XX Novel human sulfatase polypeptides useful for treating and diagnosing  
PT sulfatase-related disorders such as cerebrovascular diseases, acute  
PT meningitis, multiple sclerosis, degenerative diseases and tumor -  
XX  
PS Claim 2; Fig 15; 180pp; English.  
XX  
XX The invention provides 22438, 23553, 25278 or 26212 human sulfatase  
CC polypeptides and polynucleotides. The sulfatase genes and polypeptides  
CC are useful for treating disorders involving the brain such as  
CC cerebrovascular diseases, infections such as acute meningitis,  
CC demyelinating diseases including multiple sclerosis, degenerative  
CC diseases affecting the cerebral cortex including Alzheimer's disease  
CC and Pick disease, spinocerebellar degenerations including spinocerebellar  
CC ataxias including Friedreich ataxia, and ataxia telangiectasia,  
CC degenerative diseases affecting motor neurons including amyotrophic  
CC lateral sclerosis, inborn errors of metabolism such as leukodystrophies,  
CC toxic and acquired metabolic diseases, including vitamin deficiencies,  
CC and neurocutaneous syndromes (phakmatoses) including neurofibromatosis.  
CC The present sequence represents a human 23553 sulfatase polypeptide  
CC coding sequence.  
XX  
SQ Sequence 1800 BP; 534 A; 434 C; 423 G; 409 T; 0 other;

Query Match 78.6%; Score 1781.6; DB 22; Length 1800;  
Best Local Similarity 99.7%; Pred No. 0;  
Matches 1795; Conservative 0; Mismatches 4; Indels 1; Gaps 1;  
Qy 324 atggctccagggtgtgtgctgggggcatccgctccgccttctccacaggcctgtgtctgt 383  
Db 1 atggctccagggtgtgtgctgggggcatccgctccgccttctccacaggcctgtgtctgt 60  
Qy 384 cctggaagatgctagcaatggggcgctggcaggattctgattcctcgcctcctcaat 443  
Db 61 cctggaagatgctagcaatggggcgctggcaggattctgattcctcgcctcctcaat 120  
Qy 444 tatggttacctgtcctggggcaggcccttagaaggaggaggaagggccttactagct 503  
Db 121 tatggttacctgtcctggggcaggcccttagaaggaggaggaagggccttactagct 180  
Qy 504 caagctggagagaaactagagcccgagcaacttccactccacagcccatctcttttc 563  
Db 181 caagctggagagaaactagagcccgagcaacttccactccacagcccatctcttc 240  
Qy 564 atcctagcggatgacagggtatttagagatgtgggttaccacggatctgagattaaaaa 623  
Db 241 atcctagcggatgacagggtatttagagatgtgggttaccacggatctgagattaaaaa 300  
Qy 624 cctactctgacaagctcgtccgaaaggagtgtaactggagaaactactatgtccagct 683  
Db 301 cctactctgacaagctcgtccgaaaggagtgtaactggagaaactactatgtccagct 360  
Qy 684 atttgacacaccatccaggagtcagtttatttactggaaagtatcagatacacacccgactt 743  
Db 361 atttgacacaccatccaggagtcagtttatttactggaaagtatcagatacacacccgactt 420  
Qy 744 caacattctataagaacctaccaccccaacttttactctctctggaataatccacccta 803  
Db 421 caacattctataagaacctaccaccccaacttttactctctctggaataatccacccta 480

Qy 804 cctcagaactgaagaggttggtatattcaacgcataatggtcggaataatgacattgggt 863  
Dbb 481 cctcagaactgaagaggttggtatattcaacgcataatggtcggaataatgacattgggt 540  
Qy 864 ttttacgaaagaatgcattgcacccagagaagattgataccttttttgggtccctt 923  
Dbb 541 ttttacgaaagaatgcattgcacccagagaagattgataccttttttgggtccctt 600  
Qy 924 ttgggaagtgggtattactatacacatacaaaatgtacagtcctggatgtggttat 983  
Dbb 601 ttgggaagtgggtattactatacacatacaaaatgtacagtcctggatgtggttat 660  
Qy 984 gacttgtatgaacagcaaaatgctgctgggactatgacaaatgacatacctccacacag 1043  
Dbb 661 gacttgtatgaacagcaaaatgctgctgggactatgacaaatgacatacctccacacag 720  
Qy 1044 atgtacactcagagatcacagcaaatotttagcttccataaccccaaaagcctatat 1103  
Dbb 721 atgtacactcagagatcacagcaaatotttagcttccataaccccaaaagcctatat 780  
Qy 1104 ttatatgtcctatcaagctgttcattccaccactgcaagtcctggcaggtatttcgaa 1163  
Dbb 781 ttatatgtcctatcaagctgttcattccaccactgcaagtcctggcaggtatttcgaa 840  
Qy 1164 cactaccgatccattatacaataaacaaggagagatatgctgcattgttctcgttta 1223  
Dbb 841 cactaccgatccattatacaataaacaaggagagatatgctgcattgttctcgttta 900  
Qy 1224 gatgaacaatacaacacgtgacattggcttaaaagacttatgtttctatacaacaagc 1283  
Dbb 901 gatgaacaatacaacacgtgacattggcttaaaagacttatgtttctatacaacaagc 960  
Qy 1284 attatcatttactctcagataatggtggcagcctcagcagagagagtaactgcct 1343  
Dbb 961 attatcatttactctcagataatggtggcagcctcagcagagagagtaactgcct 1020  
Qy 1344 ctgagagtagcaagaacaataattggaagagggatccgggctgtaggttttggcat 1403  
Dbb 1021 ctgagagtagcaagaacaataattggaagagggatccgggctgtaggttttggcat 1080  
Qy 1404 agccactcttgaaacaaacaaggagagtggtgaagaacctgtgcacatacactgactgg 1463  
Dbb 1081 agccactcttgaaacaaacaaggagagtggtgaagaacctgtgtcaacatacactgactgg 1140  
Qy 1464 tacccactctcatttcaactgctgaagacagattgaggaacattcaactagatggc 1523  
Dbb 1141 tacccactctcatttcaactgctgaagacagattgaggaacattcaactagatggc 1200  
Qy 1524 tatgatattctggagaccataagtgaagggtcttcgctcaccgccagtagatatatttgc 1583  
Dbb 1201 tatgatattctggagaccataagtgaagggtcttcgctcaccgccagtagatatatttgc 1260  
Qy 1584 aacattgaccctatatacaccacaaagcaaaaatggctcctgggagcaggtatgggac 1643  
Dbb 1261 aacattgaccctatatacaccacaaagcaaaaatggctcctgggagcaggtatgggac 1320  
Qy 1644 tggaaactgcgaatccagctcagccatcagagtgacagcactggaattgcttacaggaat 1703  
Dbb 1321 tggaaactgcgaatccagctcagccatcagagtgacagcactggaattgcttacaggaat 1380  
Qy 1704 cctggctacagcagctgggtcccccctcagctctttcagcaacctgggaccgaaacgggtgg 1763  
Dbb 1381 cctggctacagcagctgggtcccccctcagctctttcagcaacctgggaccgaaacgggtgg 1440  
Qy 1764 cacaatgaacggatacaactcgtcaactggcgaagaagtgtatgcttttcaacatcacagcc 1823  
Dbb 1441 cacaatgaacggatacaactcgtcaactggcgaagaagtgtatgcttttcaacatcacagcc 1500  
Qy 1824 gaccatataagaggggtgacactatcaacaggtatccaggaatcgtaagaagctccta 1883  
Dbb 1501 gaccatataagaggggtgacactatcaacaggtatccaggaatcgtaagaagctccta 1560

Qy 1884 cgagggtctcagcttcaacaaactgcagtcgagtcgagtgatcccccaagacccc 1943  
Dbb 1561 cgagggtctcagcttcaacaaactgcagtcgagtcgagtgatcccccaagacccc 1620  
Qy 1944 agaagtacacctagctcaatggaggggtcttagggaccatggtatagagagaaccaaag 2003  
Dbb 1621 agaagtacacctagctcaatggaggggtcttagggaccatggtatagagagaaccaaag 1680  
Qy 2004 aaaaagaagcacaagcaaaaatcagctcgagaaagcaaaagcaaaagcaaaagcaaaag 2062  
Dbb 1681 aaaaagaagcacaagcaaaaatcagctcgagaaagcaaaagcaaaagcaaaagcaaaag 1740  
Qy 2063 aaaaaacagcagaagcagtcctcaggttcaactgcattcaggttacttctggataa 2122  
Dbb 1741 aaaaaacagcagaagcagtcctcaggttcaactgcattcaggttacttctggataa 1800  
RESULT 6  
AAS22726  
ID AAS22726 standard; cDNA; 1375 BP.  
XX AC AAS22726;  
XX AC AAS22726;  
XX DT 24-OCT-2001 (first entry)  
XX DE Human cDNA encoding a novel human protein #292.  
XX KW Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;  
KW immunomodulatory; cytostatic; neuroprotective; vulnerary; nootropic;  
KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;  
KW antimicrobial; antiallergic; dermatological; haemostatic; antiasthmatic;  
KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;  
KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;  
KW tissue regeneration; immune disorder.  
XX OS Homo sapiens.  
XX WO200155437-A2.  
XX PN 02-AUG-2001.  
XX PD 25-JAN-2001; 2001WO-US02623.  
XX PF 25-JAN-2000; 2000US-0491404.  
XX PR (HYSE-) HYSEQ INC.  
XX PA Tang YT, Liu C, Drmanac RT;  
XX PI WPI; 2001-451939/48.  
XX DR P-PSDB; AAU14421.  
XX PT Isolated polypeptides useful for treating anti-inflammatory diseases,  
XX PS nervous system disorders, and for regenerating bone and cartilage -  
XX Claim 1; Page 699; 894pp; English.  
CC The invention relates to polynucleotides encoding novel human  
CC proteins or their active domains. The polypeptides, polynucleotides and  
CC antibodies raised against the polypeptides are used in a method of  
CC treatment of a mammal and prevention of disorders caused by the aberrant  
CC protein expression or activity. The polypeptides can be used as  
CC molecular weight markers, food supplements, and in antibody production.  
CC The polypeptides are used to identify compounds which bind to the  
CC polypeptides. Polynucleotides of the invention are used as probes and  
CC primers, for sequencing, for chromosome or gene mapping, in the  
CC production of recombinant proteins, and in generating anti-sense DNA or  
CC RNA and in gene therapy. Polypeptides of the invention can be used to  
CC target drugs to a tumour, in assays to determine biological activity, to  
CC raise antibodies/elicit an immune response, to determine quantitative  
CC protein levels, as tissue markers, and to isolate receptors or ligands.  
CC Polypeptides of the invention may also be useful in treating platelet  
CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,



CC ligament and/or nerve tissue, wound healing, treating burns, promoting  
CC the proliferation, differentiation and survival of stem cells, as a  
CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,  
CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral  
CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or  
CC fungal infection or from autoimmunity, cancer, allergy, asthma,  
CC graft-versus-host disease, eczema, haemophilia, thrombosis,  
CC anti-inflammatory diseases, nervous system disorders, and infection.  
CC The present sequence encodes a protein of the invention.  
XX  
SQ Sequence 1375 BP; 399 A; 336 C; 360 G; 280 T; 0 other;

Query Match 39.3%; Score 890; DB 22; Length 1375;  
Best Local Similarity 100.0%; Pred. No. 5e-245;

Qy	369	caggcctgtctgtcctgctgaaagatgctagcaatgggcgctggaaggaattcttgatc	428
Db	484	caggcctgtctgtcctgaaagatgctagcaatgggcgctggaaggaattcttgatc	543
Qy	429	ctctgctctcaacttatgtttcacctgtccctggggccagcgcttagaagagagagaagaa	488
Db	544	ctctgctctcaacttatgtttcacctgtccctggggccagcgcttagaagagagagaagaa	603
Qy	489	ggggccttactagctcaagctggagagaaactagagccagccagcaaaacttccacctccag	548
Db	604	ggggccttactagctcaagctggagagaaactagagccagccagcaaaacttccacctccag	663
Qy	549	ccccatctcatcttcaatctatagcgagatgatacaggatttagagatgtgggttacacgga	608
Db	664	ccccatctcatcttcaatctatagcgagatgatacaggatttagagatgtgggttacacgga	723
Qy	609	tctgagatataaacacactactcttgacaagctcgtgccgaagagagttaaactggagaac	668
Db	724	tctgagatataaacacactactcttgacaagctcgtgccgaagagagttaaactggagaac	783
Qy	669	tactatgtccagcctatttgcacacatccaggagtcagtttattactggaaagatatcag	728
Db	784	tactatgtccagcctatttgcacacatccaggagtcagtttattactggaaagatatcag	843
Qy	729	atacacacggacttcaacattctatcataagaacctacccaaaccaactgtttacctctg	788
Db	844	atacacacggacttcaacattctatcataagaacctacccaaaccaactgtttacctctg	903
Qy	789	gacaatgccaccctactcagaaactgaaggaggttgatattcaacgcgcatatgtcggga	848
Db	904	gacaatgccaccctactcagaaactgaaggaggttgatattcaacgcgcatatgtcggga	963
Qy	849	aaatggcaacttgggtttttacagaaaaaatgcatgccccacagaagaagatttgatacc	908
Db	964	aaatggcaacttgggtttttacagaaaaaatgcatgccccacagaagaagatttgatacc	1023
Qy	909	tttttgggttcccttttgggaagtgggattactatacacactacaaatgtgcagctcct	968
Db	1024	tttttgggttcccttttgggaagtgggattactatacacactacaaatgtgcagctcct	1083
Qy	969	gggagtgtggctatgacttgtatgaaacgacaatgctgctgggactatgacaatggc	1028
Db	1084	gggagtgtggctatgacttgtatgaaacgacaatgctgctgggactatgacaatggc	1143
Qy	1029	ataatactcacacagatgtacactcagagagatagaagaaaacttagcttcccataaacc	1088
Db	1144	ataatactcacacagatgtacactcagagagatagaagaaaacttagcttcccataaacc	1203
Qy	1089	acaaagcctatatttttatattgacctataaagctgttcaattcaccaactgcaagctcct	1148
Db	1204	acaaagcctatatttttatattgacctataaagctgttcaattcaccaactgcaagctcct	1263
Qy	1149	ggcaggatatttcgaacactaccgataccattatcaacataaacagaggagagatatgtgcc	1208
Db	1264	ggcaggatatttcgaacactaccgataccattatcaacataaacagaggagagatatgtgcc	1323

Query Match		32.5%;	Score 736.2;	DB 22;	Length 1340;
Best Local Similarity		94.3%;	Pred. No. 8e-201;		
Matches 796;		Conservative 0;	Mismatches 44;	Indels 4;	Gaps 3;
QY	369	caggcctgtgtctctctggaagatctagcaatgagggcgctggcagagattctgac	428		
DB	484	caggcctgtgtctctctggaagatctagcaatgagggcgctggcagagattctgac	543		
QY	429	ctctgctctcacttattgtttacctgtccttggggccaggccttagaagaggagaa	488		
DB	544	ctctgctctcacttattgtttacctgtccttggggccaggccttagaagaggagaa	603		
QY	489	ggggccttactagctcaagctggagagaaactagagcccagcacaaacttccacctccag	548		
DB	604	ggggccttactagctcaagctggagagaaactagagcccagcacaaacttccacctccag	663		
QY	549	cccactctattttcatctcctagcgatgatcagggatttagagatgtgggtttaccacgga	608		
DB	664	cccactctattttcatctcctagcgatgatcagggatttagagatgtgggtttaccacgga	723		
QY	609	tctgagattaaacacacttactcttgacaagctcgtgcccgaaggagttaaactggagaac	668		
DB	724	tctgagattaaacacacttactcttgacaagctcgtgcccgaaggagttaaactggagaac	783		
QY	669	tactatgtccagcctatttgcacaccatccaggagtcagtttattactggaaaatcag	728		
DB	784	tactatgtccagcctatttgcacaccatccaggagtcagtttattactggaaaatcag	843		
QY	729	atacacccggacttcaacattctatcaataagacactaccacccaactgtttacctctg	788		
DB	844	atacacccggacttcaacattctatcaataagacactaccacccaactgtttacctctg	903		
QY	789	gacaatgccacctacctcagaactgaaggaggttgattatcaacgcatatgttcgga	848		
DB	904	gacaatgccacctacctcagaactgaaggaggttgattatcaacgcatatgttcgga	963		
QY	849	aaatggcacttgggtttttacagaaaagaatgcagtcgccaccacagagagattgtatcc	908		
DB	964	aaatggcacttgggtttttacagaaaagaatgcagtcgccaccacagagagattgtatcc	1023		
QY	909	tttttgggtcccttttgggaagtgggattactatacacactacaaatgtgacagt-cc	967		
DB	1024	tttttgggtcccttttgggaagtgggattactatacacactacaaatgtgacagtc	1083		
QY	968	tgggagtgtgctatgacttgtatgaaaacgacaatgctgcctgggaactatgacaatgg	1027		
DB	1084	tgggagtgtgctatgacttgtatganaacgaccatgctgcctgggaactatgacaatgg	1143		
QY	1028	catactctccacacagatgtacactcagagatgacagcaaatcttagcttcccaataccc	1087		
DB	1144	catactctccacacagatgtacactcagagatgacagcaaatcttagcttcccaataccc	1203		
QY	1088	cacaaagcctatattttatatattgctctatcaagctgttcttaccacactgcaagct-c	1146		
DB	1204	cacaaagcctataattttataattgacctcgaagcgggtctatttccactgggaggtcc	1263		
QY	1147	ctggcaggtatttcgaacact--accgatccattatcaacataaacaggagagatatgc	1204		
DB	1264	ctgggagggatttcgaacacttaccggcccttattcaacataaaggggggaggttg	1323		
QY	1205	tgcc 1208			
DB	1324	ggcc 1327			
RESULT 8					
AAA02344					
ID	AAA02344 standard; cdNA; 729 BP.				
XX					
AC	AAA02344;				
XX					
DT	19-MAY-2000 (first entry)				

XX	Human colon cancer cell line polynucleotide sequence SEQ ID NO:2335.
DE	Human; colon cancer; tumour; diagnosis; gene expression product;
XX	probe; detection; cancerous state; metastasis; identification;
KW	breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW	oestrogen receptor-negative breast cancer; lung cancer; ss.
XX	Homo sapiens.
OS	
XX	WO9958675-A2.
PN	
XX	18-NOV-1999.
PD	
XX	13-MAY-1999; 99WO-US10602.
XX	
PF	
XX	14-MAY-1998; 98US-0085426.
XX	
PR	15-MAY-1998; 98US-0085537.
PR	15-MAY-1998; 98US-0085696.
PR	21-OCT-1998; 98US-0105234.
PR	27-OCT-1998; 98US-0105877.
XX	
XX	(CHIR ) CHIRON CORP.
PA	(HYSE-) HYSEQ INC.
PA	
XX	Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI	Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI	Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI	Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX	WPI; 2000-126369/11.
DR	
XX	Polynucleotide library used to determine cancerous states of mammalian
PT	cells -
XX	
PS	Claim 1; Page 924-925; 1097pp; English.
XX	
CC	AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
CC	libraries constructed from human colon cancer cell lines. The present
CC	invention also describes a method of detecting differentially expressed
CC	genes correlated with a cancerous state of a mammalian cell, comprising
CC	detecting at least one differentially expressed gene product in a test
CC	sample derived from a cell suspected of being cancerous, where detection
CC	of the differentially expressed gene product is correlated with a
CC	cancerous state of the cell from which the test sample was detected.
CC	The polynucleotide sequences can be used in a method for detecting
CC	differentially expressed genes correlated with a cancerous state of a
CC	mammalian cell. The polynucleotides can also be used as probes for
CC	detecting and mapping related genes. They can be used in diagnosis and
CC	prognosis of diseases and disorders (e.g. identification of
CC	pre-metastatic or metastatic cancerous states, stages of cancer, or
CC	responsiveness of cancer to therapy). This is particularly for breast
CC	cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC	negative breast cancer, lung cancer, and colon cancer.
XX	
SQ	Sequence 729 BP; 232 A; 178 C; 163 G; 143 T; 13 other;
Query Match 27.6%; Score 625.2; DB 21; Length 729;	
Best Local Similarity 97.3%; Pred. No. 4.4e-169;	
Matches 644; Conservative 0; Mismatches 17; Indels 1; Gaps 1;	
QY	1407 ccacttctgaaaaaagggaacacagtgtgtaaggaacctgtgcacatcactgactggtac 1466
DB	9 cccccctttnnaaacagggaacagtggtgaaggaacctgtgcacatcactgactggtac 68
QY	1467 cccaactctcatttcactggtcgaaggacagattgatgaggacattcaactagatggctac 1526
DB	69 cccaactctcatttcactggtcgaaggacagattgatgaggacattcaactagatggctac 128
QY	1527 gatatctgggagaccataaagtggaggttcttcgctacccccagtagatatatttgcataac 1586
DB	129 gatatctgggagaccataaagtggaggttcttcgctacccccagtagatatatttgcataac 188







PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226968.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234397.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0246517.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246529.

PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-465566/50.  
DR P-PSDB; AAU23335.  
DR  
XX Novel polypeptides and polynucleotides useful for diagnosing,  
PT preventing, treating neural, immune system, muscular, reproductive,  
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous  
PT diseases  
XX

Claim 4; SEQ ID NO 431; 1180pp; English.

The present invention relates to the isolation of novel human enzyme polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences encoding them. The enzyme polypeptides of the invention may comprise the functional classes of oxidoreductases, transferases, hydrolases, lyases, isomerases or ligases. The sequences of the invention are useful in the diagnosis, treatment, prevention and/or prognosis of a wide range of disorders including hyperproliferative disorders (e.g. cancer), immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis), blood-related disorders (e.g. haemophilia), reproductive disorders (e.g. infertility) and infectious disorders (e.g. Influenza). The polynucleotides of the invention can also be used in gene therapy. AA540785-AA541684 represent cDNA sequences encoding for the novel human enzyme polypeptides of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 629 BP; 189 A; 151 C; 124 G; 162 T; 3 other;

Query Match 25.4%; Score 575.2; DB 22; Length 629;  
Best Local Similarity 97.3%; Pred. No. 9.6e-155;  
Matches 605; Conservative 0; Mismatches 15; Indels 2; Gaps 2;

Qy 577 atcagggtattagagatgtggtttaccacgagctgagattaaacacctactcttgaca 636  
Db 2 attcgccagagagatgtggtttaccacgagctgagattaaacacctactcttgaca 61  
Qy 637 agctcgctgcgaaggaggttaaaactggagaactactatgtccagcctatttgcacacat 696  
Db 62 agctcgctgcgaaggaggttaaaactggagaactactatgtccagcctatttgcacacat 121  
Qy 697 ccaggagtcagttattactggaagtatcagatcacacccggacttcaacattctatca 756  
Db 122 ccaggagtcagttattactggaagtatcagatcacacccggacttcaacattctatca 181  
Qy 757 taagacctaccaccccaactgtttacacctgtggacaatgccaccctactcagaaaactga 816  
Db 182 taagacctaccaccccaactgtttacacctgtggacaatgccaccctactcagaaaactga 241  
Qy 817 agggagttgatatcaacgcgatgtgtcggaataatggcacttgggtttttacagaaaaag 876  
Db 242 agggagttgatatcaacgcgatgtgtcggaataatggcacttgggtttttacagaaaaag 301  
Qy 877 aatgcagtcgccacagagagatttgatacctttttt-ggtcccttttgggaagtggg 935  
Db 302 aatgcagtcgccacagagagatttgataccttttttgggttcccttttgggaagtggg 361  
Qy 936 gattactatcacactacaaatgtgacagtcctggtggtgtgtggtactgactgtatgaa 995  
Db 362 gattactatcacactacaaatgtgacagtcctggtggtgtgtggtactgactgtatgaa 421  
Qy 996 aacgacaatgctgctgggactgacaaatggcagtcctggtggtgtgtggtactgactgaa 1055  
Db 422 aacgacaatgctgctgggactgacaaatggcagtcctggtggtgtgtggtactgactgaa 481  
Qy 1056 agagtacagcaaatcttagcttccataaccccaagcctatatttttatatttgc 1115  
Db 482 agagtacagcaaatcttagcttccataaccccaagcctatatttttatatttgc 541  
Qy 1116 tatcaagctgttctaccactgcaagctcctgaggtatttggacatacactacacgattcc 1175  
Db 542 tatcaagctgttctaccactgcaagctcctgaggtatttggacatacactacacgattcc 600  
Qy 1176 attatcaacataaacagagaa 1197  
Db 601 attatcaacataaacagagaa 622

RESULT 12  
AAH99356  
ID AAH99356 standard; cDNA; 562 BP.  
XX AC AAH99356;  
XX  
DT 16-OCT-2001 (first entry)  
XX  
DE Human protein encoding cDNA sequence SEQ ID NO:191.  
XX

Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
antibacterial; endocrine; cardiac; central nervous system; virucide;  
anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
antiaggregant; haemostatic; vulnery; antiulcer; osteopathic; eczema;  
dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;  
neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
genetic disease; haematopoietic disorder; platelet disorder; asthma;  
thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
allergic rhinitis; diabetes; multiple sclerosis; depression;

KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
KW neurological disorder; ss.  
XX Homo sapiens.  
OS  
PN WO200153455-A2.  
XX  
PD 26-JUL-2001.  
XX  
PF 22-DEC-2000; 200WO-US35017.  
XX  
PR 23-DEC-1999; 99US-0471275.  
PR 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT;  
XX  
XX WPI: 2001-457603/49.  
XX P-PSDB; AAM25415.  
XX  
XX Isolated human polynucleotides encoding polypeptides, useful for the  
XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection .  
PS Claim 1; Page 383; 1217pp; English.  
XX  
XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to  
XX AAM25963. The proteins can have activities based on the tissues and  
XX cells they are expressed in, such as: antiinflammatory; antirheumatic;  
XX antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;  
XX central nervous system; virucide; anti-HIV; fungicide; antimutagen;  
XX cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery;  
XX antitumor; osteopathic; dermatological; antiallergic; antiasthmatic;  
XX antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;  
XX antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
XX encoding them can be used in gene therapy, antisense therapy and vaccine  
XX production. The proteins and polynucleotides are useful for screening for  
XX agonists or antagonists of a protein and for the treatment and diagnosis  
XX of disorders associated with the activity of a protein e.g. inflammation,  
XX rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,  
XX neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
XX infections, autoimmunity, genetic diseases, haematopoietic disorders,  
XX anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,  
XX osteoporosis, severe combined immunodeficiency, eczema, allergic  
XX rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
XX Alzheimer's disease, Parkinson's disease, neurodegenerative and  
XX neurological disorders.

Sequence 562 BP; 170 A; 132 C; 105 G; 155 T; 0 other;

Query Match 23.8%; Score 538.4; DB 22; Length 562;  
Best Local Similarity 99.8%; Pred. No. 3.4e-144;  
Matches 539; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 719 aaggtatcagatcacacccgagcttcaacattctatcatgaagactcaccacccaactg 778  
Db 21 aaggtatcagatcacacccgagcttcaacattctatcatgaagactcaccacccaactg 80  
Qy 779 ttacctctggacaatgccaccctaccctcagaacactgaagggttgatatttcaacgca 838  
Db 81 ttacctctggacaatgccaccctaccctcagaacactgaagggttgatatttcaacgca 140  
Qy 839 tatggtcggaaaaatggcacttgggtttttacagaaaaaagaatgcgtccaccagagaagg 898  
Db 141 tatggtcggaaaaatggcacttgggtttttacagaaaaaagaatgcgtccaccagagaagg 200  
Qy 899 attgataccttttttgggttcccttttgggaagggttggtactactacacactacaagt 958  
Db 201 attgataccttttttgggttcccttttgggaagggttggtactactacacactacaagt 260  
Qy 959 tgacagtcctgggatgtgtggctatgactgtatgaaacacgaactgtcgtcgggacta 1018







PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;  
 PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;  
 PI Lamson G, Drmanac R, Krkenjakov R, Dickson M, Drmanac S, Labat I;  
 PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;  
 XX WPI; 2000-126369/11.  
 DR  
 XX  
 PI Polynucleotide library used to determine cancerous states of mammalian  
 PT cells -  
 XX  
 PS Claim 1; Page 938; 1097pp; English.  
 XX  
 CC AAA00010 to AAA02716 represent polynucleotides isolated from cDNA  
 CC libraries constructed from human colon cancer cell lines. The present  
 CC invention also describes a method of detecting differentially expressed  
 CC genes correlated with a cancerous state of a mammalian cell, comprising  
 CC detecting at least one differentially expressed gene product in a test  
 CC sample derived from a cell suspected of being cancerous, where detection  
 CC of the differentially expressed gene product is correlated with a  
 CC cancerous state of the cell from which the test sample was derived.  
 CC The polynucleotide sequences can be used in a method for detecting  
 CC differentially expressed genes correlated with a cancerous state of a  
 CC mammalian cell. The polynucleotides can also be used as probes for  
 CC detecting and mapping related genes. They can be used in diagnosis and  
 CC prognosis of diseases and disorders (e.g. identification of  
 CC pre-metastatic or metastatic cancerous states, stages of cancer, or  
 CC responsiveness of cancer to therapy). This is particularly for breast  
 CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-  
 CC negative breast cancer, lung cancer, and colon cancer.  
 XX  
 SQ Sequence 1082 BP; 230 A; 182 C; 155 G; 190 T; 325 other;

Query Match 6.2%; Score 139.6; DB 21; Length 1082;  
 Best Local Similarity 93.08; Pred. No. 2.1e-29;  
 Matches 187; Conservative 0; Mismatches 9; Indels 5; Gaps 4;  
 QY 1493 acagattgagagacattcaacta-gatgctgtatgatctggagaccataagtggag 1551  
 DB 268 acagattgagagacattcaactaactaggatgctgtatgatctggagaccataagtggag 327  
 QY 1552 gtcttcgtcaaccgagtagatatttgcataaaca-ttgaccccatatacaccaaggca 1610  
 DB 328 gtcttcgtcaaccgagtagatatttgcataaaca-ttgaccccatatacaccaaggca 387  
 QY 1611 aaaaatggctctggcagcagcgtat-ggagctgtgaacactgcaatccagtcagccat 1669  
 DB 388 aaaaatggctctggcagcagcgtat-ggagctgtgaacactgcaatccagtcagcagc 447  
 QY 1670 --cagagtgcagcactggaaa 1688  
 DB 448 tcaagaggcagcactggaaa 468

RESULT 15  
 ABL12859  
 ID ABL12859 standard; cDNA; 1740 BP.  
 XX  
 AC ABL12859;  
 AC  
 XX 26-MAR-2002 (first entry)  
 DT  
 XX  
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 33059.  
 XX  
 KW Drosophila: developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ss.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN W0200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX

PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 PI  
 XX WPI: 2001-656860/75.  
 DR P-PSDB; ABB68756.  
 DR  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 XX Claim 1; SEQ ID NO 33059; 2lpp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 1740 BP; 469 A; 420 C; 441 G; 410 T; 0 other;  
 Query Match 5.7%; Score 129; DB 23; Length 1740;  
 Best Local Similarity 48.5%; Pred. No. 3.1e-26;  
 Matches 502; Conservative 0; Mismatches 500; Indels 33; Gaps 4;  
 QY 547 agcccatctcatcttccatcctagcggatgatcaggatgttagagatgtgggtaccacg 606  
 DB 80 aaccaataatagtataataactaatactgatcagatggcagatgcagtgtttcag 139  
 QY 607 gatc---tgagattaaaaacactactcttgacaagctcgtgcggaaggaggttaactgg 663  
 DB 140 gctcgaaatcagatccttacaccgaataagatgctttggcctacaatggtatcctgttga 199  
 QY 664 agaactactatgtccagcctatttgcacacacatcagcaggtcagttattactggaagt 723  
 DB 200 atagcatattgtccgaaacctatgcacccatccagagctactctgctcactggaat 259  
 QY 724 atcagatacacacccggacttcaacttctcataaagacctaccaccaactgttttac 783  
 DB 260 atccatacacacagcagcagcactttgtattatcacagatgagcgtgggccccttc 319  
 QY 784 ctctggacaatgccaccctactcagaactgaaggaggttgatattcaacgcataagg 843  
 DB 320 ctgagcgagaacgtcttatgcgcaaatcttccggaggtgcgggtactcgacccattgg 379  
 QY 844 tcgaaaaatgcacttgggtttttacagaaaagatgatccccaccagaagagatttg 903  
 DB 380 tggcaagtggcacttgggtctcggcgaaggatctcacccaacaatgctggattcgc 439  
 QY 904 atacctttttggttcccttttgggaagtggggtattactatcacactacaaatgtgaca 963  
 DB 440 accatcatttggctactacaatggctacattgattactacagaccatcaggtgcgaatgc 499  
 QY 964 gtcttgggaatgtgtggtatgactgtatgataaaacgacaatgctg---cctgggactatg 1020  
 DB 500 tcgataggaaactactcgggtgactggtattcccgggggtctgagcgtgccccgaag 559  
 QY 1021 acaatggcatatactccacacagatgtacactcagagatcagacaaattcttagttccc 1080  
 DB 560 cgaatggcacttaacgccacgaagccttacttctggaagcaaaaggataattgagcaac 619



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 17:15:04 ; Search time 69.6 seconds  
(without alignments)  
759.327 Million cell updates/sec

Title: US-09-495-823-7  
Perfect score: 3012  
Sequence: 1 MAPRCAGHPPPPSPQACVC.....VPRYPKPPSRNPLNGGV 550

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_71.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1424.5	47.3	533	1 KJHUAB	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
2	1414.5	47.0	535	1 A44475	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
3	1333	44.3	473	2 I54210	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
4	472	15.7	521	2 H87394	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
5	454	15.1	497	2 AF0306	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
6	429	14.2	536	2 D83622	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
7	422.5	14.0	522	1 KJHUG6	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
8	416.5	13.8	567	2 A37362	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
9	413	13.7	583	1 KJHUAC	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
10	412.5	13.7	533	2 S69336	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
11	407.5	13.5	589	2 I37187	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
12	399	13.2	593	2 I37186	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
13	398	13.2	551	2 C91220	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
14	398	13.2	551	2 E86066	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
15	395.5	13.1	537	2 F95993	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
16	394	13.1	551	2 S30691	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
17	393	13.0	555	2 D95377	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
18	393	13.0	577	2 T45348	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
19	378	12.5	507	1 KJHUAA	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
20	364	12.1	506	2 A54190	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
21	357.5	11.9	551	2 S07089	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
22	356	11.8	464	2 B35159	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
23	349.5	11.6	551	2 AF0103	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
24	349.5	11.6	551	2 S01793	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
25	344.5	11.4	303	2 T15835	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
26	336	11.2	545	2 F95325	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
27	325	10.8	571	2 B85726	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
28	324	10.8	571	2 E64903	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human
29	324	10.8	571	2 G90891	N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human

30	319.5	10.6	787	2	B70643	probable sulfatase
31	314.5	10.4	557	2	AD0370	probable sulfatase
32	272	9.0	495	2	AF3523	choline-sulfatase
33	257.5	8.5	517	2	AB0102	probable sulfatase
34	256	8.5	452	2	T29618	hypothetical prote
35	248.5	8.3	552	1	KJHUGU	N-acetylglucosamin
36	247	8.2	649	2	S43229	arylsulfatase (EC
37	243.5	8.1	970	2	E70533	probable sulfatase
38	243	8.1	497	2	C91206	probable sulfatase
39	243	8.1	497	2	E86052	probable sulfatase
40	241	8.0	497	2	G65169	probable sulfatase
41	221.5	7.4	503	2	E83642	choline sulfatase
42	219.5	7.3	709	2	T16584	hypothetical prote
43	213.5	7.1	787	2	B70535	probable sulfatase
44	212	7.0	538	2	F83354	probable sulfatase
45	205.5	6.8	550	1	KJHUID	iduronate-2-sulfat

ALIGNMENTS

RESULT 1

KJHUAB

N-acetyl-galactosamin-4-sulfatase (EC 3.1.1.6.12) precursor [validated] - human

N;Alternate names: arylsulfatase B (ASB); chondroitinase; chondroitinsulfatase; G4S;

C;Species: Homo sapiens (man)

C;Date: 31-Dec-1993 #sequence\_revision 27-Oct-1995 #text\_change 08-Dec-2000

C;Accession: S35990; S33307; A35078; A45659; A42449; B42449; I54217; A56865

R;Peters, C.W.B.

Submitted to the EMBL Data Library, March 1993

A;Reference number: S35990

A;Accession: S35990

A;Molecule type: DNA

A;Residues: 1-533 <P>

A;Cross-references: EMBL:X72735; NID:g289009; PIDN:CAA51272.1; PID:g825628

R;Modaresi, S.; Rupp, K.; von Figura, K.; Peters, C.

Biol. Chem. Hoppe-Seyler 374, 327-335, 1993

A;Title: Purification of the human arylsulfatase B gene.

A;Refer

A;Access

A;Molec

A;Resid

A;Cross

A;Note:

R;Peters

J. Biol

A;Title

A;Refer

A;Access

A;Molec

A;Resid

A;Cross

A;Note:

R;Litjel

Biochem

A;Title

A;Refer

A;Access

A;Molec

A;Resid

A;Cross

A;Note:

R;Jin,

Am. J.

A;Title: Mucopolysaccharidosis type VI: identification of three mutations in the aryl

city.

A;Reference number: A42449; MUID:92197625

A;Accession: A42449

A;Molecule type: mRNA

EMBL:X72738; EMBL:X72739;

Isidor, M.; Vingron, M.; M

cloning and expression of

; PID:gl79077

equencing

Hopwood, J.J.

maturation and isolation o

; PID:g236698

NCBIP:57778)

the mature protein, were d

ewhere between residue 450

102 (5) >451



Db 39 -----AGAG-----ADRPPLHVFVLADLGNWVSHGNSIRTRPHLDELAAAGVLLD 85

Qy 115 NYVQPICTSRSFITGKYQIHTGLQHSIRTPQPNCLPLDNATLPQKLKEVGYSTHMV 174

Db 86 NYVQPICTSRSQLLGRYQIHTGLQHIWPCQPSVPLDEKLLPQLKEAGYTHMV 145

Qy 175 GKWLHGIRKECMPTRRGFDFTFGSLGSDYTHYKC----DSPGM--CGYDLYENDNAA 229

Db 146 GKWLHGIRKECLPTRRGFDFTFGYLLGSEDTYSHERCALIDSLNVTFCALDFRDGEQVA 205

Qy 230 WYDNGIYSTQMTYTORVQOITLASHNPKPIFLXAYQAVHSPLOAPGPRYFHYRSIININ 289

Db 206 TGYKN-MYSNIFTERATALTSHPPKPLFLYLALQSHVPEYLPKDYFIDOKN 264

Qy 290 RRYAAMLSCLDEAINNVTLAKTYGYNNIIISYSDNGQPTAGSNMPLRGSKGTW 349

Db 265 RHYVAGVMSLDEAVGNVTAALKSHGLWNNTVFIFSTDNGQTLAGGNMPLRGKWSLW 324

Qy 350 EGGIRAVGFVHSPLLKNGKTCVKEPVHITDWTYPTLISLAEGQDDEDIOLGDIWETISE 409

Db 325 EGGIRGVGFVASPLLKQKGVKNRELIIHSDWLEPTLVKARGSTKGTPLDGFVWKTISE 384

Qy 410 GLRSPRVIDLHNIDPIY-----TRAKNGSWAAGYGNWNTAIOASAIRVQHWKLLT 458

Db 385 GSPSPRELHNIDPNFVDISPCPKSLAPAKDDSSHPAYLAFNTLSHAIRHGNWKLTT 444

Qy 459 GNGYSOWPVPQSFNSLGNRWHNERITSS--TGKSYWLFNITADPYERVDLSNRYPGIV 516

Db 445 GYPGCGCWFP-----PSPYNDISAIPSSDPPTKTLWPFDIDQDPEERHDLSDRYPHIV 497

Qy 517 KLLRLRSQNKTAVPVRYPPKDPSPRNPLNG 548

Db 498 EQLLSRLQFYHKKHVSVPVHFPAQDPRCDPKGTG 529

RESULT 3

I54210

N-acetylglactosamine-4-sulfatase (EC 3.1.6.12) precursor - rat (fragment)

N:Alternate names: arylsulfatase (EC 3.1.6.1) B [misidentification]

C:Species: Rattus norvegicus (Norway rat)

C:Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 02-Jun-2000

C:Accession: I54210

R:Kunieda, T.

Genomics 29, 582-587, 1995

A:Title: Mucopolysaccharidosis type VI in rats: Isolation of cDNAs encoding arylsulfatases

A:Reference number: I54210; MUID:96121368

A:Accession: I54210

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-473 <RES>

A:Cross-references: GB:D49434; NID:g1065603; PIDN:BAA08412.1; PID:g1089794

C:Genetics:

A:Gene: ARSB

C:Superfamily: animal sulfatase

C:Keywords: sulfuric ester hydrolase

F:31/Modified site: 3-oxoalanine (Cys) #status predicted

Query Match 44.3%; Score 1333; DB 2; Length 473;

Best Local Similarity 53.9%; Pred. No. 2.6e-94;

Matches 255; Conservative 70; Mismatches 122; Indels 26; Gaps 5;

Qy 94 HGSIKTPTLDKLAAGVKGLENNYVQPICTPSRSQFITGKYQIHTGLQHSIRTPQPNCL 153

Db 3 HGSVIRPHLDALAAGVLDNYYVQPLCTPSRSQLTGRYQHGLQHYLIMTCQNCV 62

Qy 154 PLDNATLPQKLKEVGYSTHMVGVKWHLGFYRKECMPTRRGFDFTFGSLGSGDYTHYTH 209

Db 63 PLDEKLLPQLKDGSSSTMVGVKWHLGMRYKECLPTRRGFDFTFGYLLGSDYTHYTH 122

Qy 210 -YKCDSPGMCYDLYENDNAAWDYDNGIYSTQMTYTORVQOILASHNPKPIFLYLIAQAV 268

Db 123 PIECLNTRCALDRDGEAPKAYTD-IYSTNIETKTRATLIANHPPEKPLFLYLAFQSV 181

Qy 269 HSPLQAPGRYFHYRSIININRRRYAAMLSCLDEAINNVTALKTYGYNNIIISYSDN 328

Db 182 HDPLQVPEEYMEPYDFIQDKHRRIRYAGVMSILLDEAVGNVTALKSRGLWNVTLIFSTDN 241

Qy 329 GGQPTAGGSNWPPLRGSKGTWEGIRAVGFVHSPPLKNGKTCVKEPVHITDWTPTLISLA 388

Db 242 GGTRSGNWNPLRGSKGTWEGIRAVGFVHSPPLKNGKTCVKEPVHITDWTPTLISLA 301

Qy 389 EGQIDEDIQDGDYDIWETISEGLRSRVIDLHNIDPIY-----TKAKNGSWAAG 437

Db 302 GGSTHGTPKLDGFDVWETISEGSPSPRVELLNIDDPDFDGLPCPGKNTYTPKNDSEPLE 361

Qy 438 YGIWNTAIOASAIRVQHWKLLTGNPGYSDWVPPQSFNSL--PNRWHNERITSSYTKSVWL 495

Db 362 HSAFNTSIHAGIRYKKNWKLTYGPGCGYWFPPSPQSNISEVPS-----VDSPTKTLWL 414

Qy 496 FNITADPYERVDLSNRYPGIVKLLRLRSQNKTAVPVRYPPKDPSPRNPLNG 548

Db 415 FDIRNDPEERHDSREHPHIVQNLISRLQYHHSVSPYFPPLDPRCDPKGTG 467

RESULT 4

H87394

sulfatase family protein [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C:Accession: H87394

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko, n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: H87394

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-521 <STO>

A:Cross-references: GB:AE005673; NID:g13422494; PIDN:AAK23156.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC1172

Query Match 15.7%; Score 472; DB 2; Length 521;

Best Local Similarity 26.6%; Pred. No. 3.2e-38;

Matches 142; Conservative 69; Mismatches 193; Indels 130; Gaps 16;

Qy 68 PSTTSTSQPHLIFILADDDQGFDRV-----GYHSEIKTPTLDKLAAGVKGLEN-YYVQPI 121

Db 25 PPADGQRPPIVIFILADDDGYNITLNGGGVAGGTVPPTPAIDSTAKEGVTFANGYSGNAT 84

Qy 122 CTPSRQFITGKYQIHTGLQHS-----IIRTPQPN----- 151

Db 85 CAPSRAAIMTGRYATRFGEFTPTPVAFSRVVGHGAGDPLHPSRFNQAEVKNMFKDENVL 144

Qy 152 CLPLDNATLPQKLKEVGYSTHMVGVKWHLGFYRKECMPTRRGFDFTFGSLGSG-----GD 205

Db 145 AVPAAEVTIAEALTKGYHTHLGKWHLGGV-KGSRPEDQGFDESGLGFMAGALFAPVGD 203

Qy 206 YYTHYKCDSPGMCYDLYENDNAAWD-----YDNGIYSTQMTY 244

Db 204 -----PGV-----ESRQDWDPIDKFLWGAAPFAVQFNGGKLFNFSHYMTDYLTD 248

Qy 245 RVQOILASHNPKPIFLYLIAQAVHSPLOAPGPRYFHYRSIININRRRYAAMLSCLDEAI 304

Db 249 EAVRAIDA-KNRPFEMFLAYNAVHTLQAPKADYDALSHIKDHMRVYAAWNRNLDNRV 307

Qy 305 NNVTALKTYGYNNIIISYSDNGQPTAG--GSMNPLRGSKGTWEGGIRAVGFVHSP 362

Db 308 GKVLQALKDRGLDDNTLVIFTSNDGGANYICLPDINKPYRGWKATFFEGGKLVFLLRWP 367

Qy 363 LLKKNKGTCVKEPVHITDWTPTLISLAEGQDDEDIOLGDIWETISEGLRSRVIDLHNI 422



Db 368 EQLPAGAVYRSPGVHVDIFATAAGAAGAPTPKDRVLDCVDL-----VFFVKQSSG 418  
Qy 423 DPIYTKAKNGSWAAGYGIWNTAIQSAIRVQHWKLLTGNPGYSDWVPPQSFNSLGNRWHN 482  
Db 419 DP--HKAI-----FWRSGGYKTVLAGWKQVAK-----PNK---- 449  
Qy 483 ERITSSGKSWLNFNITADPYRVDLSNRYPGVIGVKLLRLRLSQFNKTAAPVRYP 536  
Db 450 -----TWLPDLSDTPERQELSKARPEKLEKREMQAIMAQLDQGMKKTWP 493  
RESULT 5  
AF0506  
C:Species: *Salmonella enterica* subsp. *enterica* serovar Typhi  
A:Note: This species has also been called *Salmonella typhi*  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 09-Nov-2001  
C:Accession: AF0506  
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; Farrar, M.; Mouton, S.; O'Garra, P.; S.; Mouton, S.; O'Garra, P.  
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi  
A:Reference number: AB0502; PMID:11677608  
A:Accession: AF0506  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-497 <PAR>  
A:Cross-references: GB:AL513382; PIDN:CAD01190.1; PID:g16501319; GSPDB:GN00176  
C:Genetics:  
A:Gene: STY0042

Query Match 15.18; Score 454; DB 2; Length 497;  
Best Local Similarity 26.38; Pred. No. 7.1e-27;  
Matches 138; Conservative 77; Mismatches 189; Indels 120; Gaps 16;  
Qy 62 AGEKLEPSTTTS-QPHLIFILADDOGRDVGHYGSEI-KPTLDKLAAGVKLENYV- 118  
Db 16 AGCVLSAQAATAKRPNLVIILADDLGVLATYGHRIKVTNPIDKLAQEGVKETDIYAP 75  
Qy 119 QPCTPSSQITGKYQIHTGLQHSIIRPTQPNCLPDNATLPQKLEKVGYSYTHWVKWH 178  
Db 76 APLCSPSRAGLLTGMPRTGIR-SWIPEGKDVAGLGRNELTIANLLKQOQYDTAMMGKLI 134  
Qy 179 L--GFKYKCEMPTRRGDTFFGSLGSDYTHYKCDSPGCMGYDLYENDNA----- 228  
Db 135 LNAAGDRDTPQAKDMGFDY---TLVNPAGVTDATL-----DNAKERPRYG 178  
Qy 229 -----AWDYDNGY-----STQMYTORVOQILASHNPTKPIFLYAYQAVHSPLOAPG 276  
Db 179 VVHPTGW-IRNGHIGRADKMSGEFVSSEVYNWLDNKKDDNPFFLYVAFTEVHSPLASPK 237  
Qy 277 RYFHYRSIIINRR-----YAMLSCLEAINNVTLAKTYGFY 317  
Db 238 KYLDMYSQYMTDYQKHQDPLFGYDWDKPRGTGTYEYANISYMDYEQGVKLDKIKAMGEE 297  
Qy 318 NSIIITYSSDNGG-----OPTAGSNWPLRSGTYGEGGIRAGVFY-----HSPLL 364  
Db 298 DNTIVFTSDNGPVYTRARKYVELKAGETGLGRKDNLEGGGIRVPAITKYGKHIP-- 355  
Qy 365 KNGTYCKEYVHITDWPYTLISLAEGQIDEDQLDGDYDIWETISEGLRSPRVLDLHNIDP 424  
Db 356 --QGMVTDTPVYGLDWLPTLANMMDFKLPTDRTYDQGSVLPLDKTLKQKPLIFGIDM 413  
Qy 425 IYTKAKNGSWAAGYGIWNTAIQSAIRVQHWKLLTGNPGYSDWVPPQSFNSLGNRWHNER 484  
Db 414 PFQDDPTDEW-----AIRDGDWKN----- 432  
Qy 485 ITSSTGKSWLNFNITADPYRVDLSNRYPGVIGVKLLRLRLSQFNK 528  
Db 433 IIDRQNKPKYILNLTDRFELNIGKQPOKQYKGLFKYK 476

RESULT 6  
D83622  
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen  
C:Species: *Pseudomonas aeruginosa*  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: D83622  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Adams, S.; Ryan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L.; Lory, S.; Olson, M.V.  
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen  
A:Reference number: AB2950; MUID:20437337  
A:Accession: D83622  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-536 <STO>  
A:Cross-references: GB:AE004456; GB:AE004091; NID:g9946013; PIDN:AAG03573.1; GSPDB:GN00176  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: *atsA*; PA0183  
C:Superfamily: animal sulfatase

Query Match 14.28; Score 429; DB 2; Length 536;  
Best Local Similarity 24.98; Pred. No. 6.5e-25;  
Matches 144; Conservative 71; Mismatches 130; Indels 174; Gaps 20;  
Qy 73 TSQPHILFILADDOGRDVGHYGSEI-KPTLDKLAAGVKLENYVQPIPTSPRSQFTIG 132  
Db 2 SKRPNFLVIVADDLGFSDIGAFGEIATPNDALAIAGLITDFTASTCSPTRSMLL-- 59  
Qy 133 KYQIHTGLQHSII-----RPTQPNCLPDNATLPQKLEKVGYSYTHWVKW 177  
Db 60 -----TGTGDHIAIGTMAEALTEPECKPGYEGHNRVVALPELLREAGYQTLMAKW 114  
Qy 178 HLGFYRKECMPTRRGDTFFGSLGSDYTHYKCDSPGCMGYD-----LYEN 225  
Db 115 HLGL-KPEQTPHARGFERSFLLFGAAN---HYGFEPP---YDESTPRILKGTALYVE 166  
Qy 226 DNAAMD-YDNGIYSTQMYTORVOQILASHNPTKPIFLYAYQAVHSPLOAPGRYFEHYRS 284  
Db 167 DERYLDTLPEGYSSDAFGDKLLQYKQDQSRPFPFAYLPPFSAPHWLPQAPREIVEKYRG 226  
Qy 285 -----IININR-----RRYAAAM 296  
Db 227 RYDAGPEALROERLARLKEGLVLEADVEAHPVLAALTREWEALEDEERAKSARAMEVYAA 286  
Qy 297 LSCLEAINNVTLAKTYGFYNNIIYSSDNGG-----PTAG----- 335  
Db 287 VERMDWNGRVVDYLRROGELDNFTVLFMSDNGAEGALLEAFKPGPDLLGLDRHYDNS 346  
Qy 336 -----GSNW-----PLRSGKTYWEGGIRAGVFHSPVPLKNGTKVCKEYVH 376  
Db 347 LENIGRANSYVYVYVPRWAQAATAPSRLYKFTTQGGIRVPALVRYPRLSRQCAISHAFAT 406  
Qy 377 ITDWPYTLISLAEGQIDEDQLDGDYDIWETISE---GLRSPRVLDLHNIDPIYTKAKNGS 433  
Db 407 VMDVPTLLDLA-GVRHPGKRWGRGRIEAPGRSWLGSWLSGSETAAHDENTV----- 457  
Qy 434 WAAGYGIWNTAIQSAIRVQHWKLLTGNPGYSDWVPPQSFNSLGNRWHNERITSTGKSV 493  
Db 458 --TG---WELFGMRAIROGDMKAV-----YLP---APVGATWQ----- 488  
Qy 494 WLFNITADPYRVDLSNRYPGVIGVKLLRLRLSQFNKTAAPVRYP 531  
Db 489 -LYDLARDPGEIHLADSPQKGLAEILIEHWKRYVSETGV 526

RESULT 7  
KJHUG6  
N-acetylglactosamine-6-sulfatase (EC 3.1.6.4) precursor [validated] - human

N:Alternate names: chondroitinase; chondroitinsulfatase; galactose-6-sulfate sulfatase;  
C:Species: Homo sapiens (man)  
C:Date: 31-Mar-1992 #sequence\_revision 27-Oct-1995 #text\_change 08-Dec-2000  
C:Accession: JQ1299; PQ0242; I37406  
R:Tomatsuo, S.; Fukuda, S.; Masue, M.; Sukegawa, K.; Fukao, T.; Yamagishi, A.; Hori, T.;  
ashi, Y.; Orit, T.  
Biochem. Biophys. Res. Commun. 181, 677-683, 1991  
A:Title: Morquio disease: isolation, characterization and expression of full-length cDNA  
A:Reference number: JQ1299; MUID:92095973  
A:Accession: JQ1299  
A:Molecule type: mRNA  
A:Residues: 1-522 <TOM>  
A:Experimental source: placenta  
A:Accession: PQ0242  
A:Molecule type: protein  
A:Residues: 27-42, 'X', 44-54; 175-183, 'X', 185-192; 201-218; 243-265; 311-324, 'XX', 327-336; 377-  
R:Morris, C.P.; Guo, X.H.; Apostolou, S.; Hopwood, J.J.; Scott, H.S.  
Genomics 22, 652-654, 1994  
A:Title: Morquio A syndrome: cloning, sequence, and structure of the human N-acetylgalac-  
A:Reference number: I37406; MUID:95095267  
A:Accession: I37406  
A:Status: preliminary; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-522 <RES>  
A:Cross-references: EMBL:U06088; NID:g507365; PIDN:AAC51350.1; PID:g618426  
A:Note: nucleotide sequence not complete  
C:Genetics:  
A:Gene: GDB:GALNS  
A:Cross-references: GDB:129085; OMIM:253000  
A:Map position: 16q24-16q24  
A:Introns: 40/3; 82/1; 107/1; 141/2; 189/2; 253/2; 300/1; 334/3; 380/2; 414/3; 459/2  
A:Note: defects in this gene can cause mucopolysaccharidosis type IV A, Morquio disease  
C:Function:  
A:Description: hydrolyzes N-acetylgalactosamine-6-sulfate units in chondroitin sulfate  
C:Superfamily: animal sulfatase  
C:Keywords: glycoprotein; sulfuric ester hydrolase  
F:1-26/Domain: signal sequence #status predicted <SIG>  
F:27-522/Product: N-acetylgalactosamine-6-sulfatase #status experimental <MAT>  
F:79/Modified site: 3-oxoalanine (Cys) #status predicted  
F:204,423/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 14.08; Score 422.5; DB 1; Length 522;  
Best Local Similarity 25.4%; Pred. No. 2e-24;  
Matches 142; Conservative 84; Mismatches 225; Indels 109; Gaps 20;

QY 24 MAMGALAGFWILCLLYGYLSNGQALLEEELGALLAQAGEKLEPSTSTSQPHLIFILA 83  
DB 1 MAAVATRWQQLL-----VLSAAG---MGASGAPOPPNILLLLM 38

QY 84 DDQGFDRDVGYHGSEIK-TPTLDKLAAGVKNLNY-VQPICTPSRSQFITGKYOIHGTL- 140  
DB 39 DDMGWDGLVYGEPSRETPLNDRAAEGLLFPNYSANPLCSPSRAALLTGRLPIRNGFY 98

QY 141 -----QHSIIIRPTQ-PNCLPLDNATLPOKLEKGVGSHYHWGKWHLGFYRKCEMPTRRGFD 194  
DB 99 TTNAHARNAVTPQIEIGVDPDSEQLLPELLKKAGYVSKIVGKWHLG-HRPQFHLKHGFD 157

QY 195 TFFGSLGSGDYTHYKCDSPGCMGYDLYENDNA-----WDYDNGIYS----- 238  
DB 158 EWFG-----SPN-CHFGDQNKARNPIPVYRDWMVMGVRYEEFFPNLKTG 201

QY 239 ----TQMYTORVQOILASHNPTKPIELYIAVOAHSPLQAPGRYFEHYRSTININRRYA 294  
DB 202 EANTQIYLQALDFIKRAHRHFFELYWADVATHAPVYASKP-----LGTSGRIGY 255

QY 295 AMLSCLDEAINNVTLAKTKYGFYNNSTIIYSSDNG-----GQPTAGGSNWPLRGSKGYWE 350  
DB 256 DAVREIDDSIGKILELLQDLHVADNTVFVFTSDNGAALISAPEGGSGNPFLCCKQTFE 315

QY 351 GGIRAVGFVHSPLKNGKTVCKEVPVHTDMVPTLISAEQDIEDIQLDGYDIWETISEG 410  
DB 316 GGMREPALWMPGHVTAQGVSHQJLSIMDLFTTSLAAGLTTPPSDRAIDGLNLLPTLQOG 375

```

Qy 411 LRSPRVILHNDIPYKAKNGSWAAGYGIWNTAIQSAIRVQHWKLLTGNGPYSDWVPPQ 4770
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 376 RLMDRPIFYRGDTLMA-ATLGQKHAHEFTWTS-----WENFRQG---IDFCPGQ 422
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 471 SFSNLGPNRNRHETSTSGKSVMLFNITADPYERVDL---SNRYPGIVKLLRLRSQFN 527
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 423 NVSGV---TTHN---LEDHTKLPIFIHLGRDGPGRFPLSFASAEYQEALSRIYSVQOHQ 476
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Qy 528 KTAVPVRYPPKDRSPRNLN 547
      : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 477 EALVP-----AQPQLN 487
      : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 8
A37362
arylsulfatase (EC 3.1.6.1) precursor - sea urchin (Strongylocentrotus purpuratus)
C:Species: Strongylocentrotus purpuratus (purple urchin)
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 24-Nov-1999
C:Accession: A37362
R:Yang, Q.; Angerer, L.M.; Angerer, R.C.
Dev Biol 135, 53-65, 1989
A:Title: Structure and tissue-specific developmental expression of a sea urchin aryls
A:Reference number: A37362; MUID:89357267
A:Accession: A37362
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-567 <YAN>
A:Cross-references: GB:M28404; GB:M25815; NID:g161440; PID:g161441
C:Superfamily: animal sulfatase
C:Keywords: sulfuric ester hydrolase
P:115/Modified site: 3-oxoalanine (Cys) #status predicted

```

Query Match	13.8%;	Score	416.5;	DB	2;	Length	567;
Best Local Similarity	29.7%;	Pred. No.	6.4e-24;				
Matches	120;	Conservative	66;	Mismatches	175;	Indels	43;
Gaps	14;						

  

Qy	58	LLAQAQKLEPSTSTSPH	LIFLADDDGPRDVG	YHGSEIKTP-TL	DKLAAEGVKL-EN	115
Db	53	LLGTGQ----	HRTAMKPNVILLAD	DGMVGDSLVS	YGHPTQEPFI	DQMANQGLRTQG
Qy	116	YVVOICHPRSSQITG	KYQIHTGL--OHS	TIIRTPQNCILP	DNATLPQKLKEV	GYSTHM
Db	109	YSGDSVCTPSSRA	IVTGRQIRTVGYE	ERIFLPWT	TGLPLYEV	YETIAEMKAGAYTGM
Qy	174	VKKHHLGFRKEC----	MPTRRGFDTF	FFGSLGSGDY	YTHYKCDSP	CMCGYDLYENDNA
Db	169	VKKHHLGHNENSS	SDGAHLNANRGFD-FV	GHNLFPGN---SWR	CDTGL--HQD	FPDPTNA
Qy	229	AWDYDNGIYSTQMY	TQR-VOOIL-----	ASHNPTKPIFL	YATAYQAVHSP	LQAPGRYFE
Db	223	CFLYNSTSAQPFQ	HKGTLTOLLRDDT	VGFTEDMNVK	PFMYVFAHMTSL-----	FS
Qy	281	HYRSIINRRRYA	AMLSCLEDAENNV	TALKTYGFYNN	SIITYSSDNG-----	GOPTA
Db	277	SDDFCTSRRGY	GDNREMDQAEQ	IVTTLVND	IDDNTVIFTS	DGHPRHYCGE---
Qy	335	GSGNPLRGSGT	YVWEGIRAGV	FVHSPHLLKN	KTKCTCKPEPH	ITDWPYTLISLAEGOIDE
Db	334	GGDANVFSGK	QWSEGGRPIY	IVYPGTISP	G-VSHEIVTSM	DIATAVNLGGSOLPT
Qy	395	DIQLDGYDIWET	I	SEGURSPVD	ILHNIDPIY	YTKAKNGSNAAGY
Db	393	DRYDGGKCLK	SVLLEGASSPH	DDPFYICKD	PLMAVRVGYKAHF	436

RESULT 9  
K3HUAC  
steryl-sulfatase (EC 3.1.6.2) precursor - human  
N:Alternate names: arylsulfatase C; steroid sulfatase (STS); steryl-sulfate sulfohydrolase  
C:Species: Homo sapiens (man)  
C:Date: 21-May-1990 #sequence revision 27-Oct-1995 #text change 11-May-2000

C;Accession: A32641; A57116; A25961; S05415; S05423; I52800; I65619  
R;Stein, C.; Hillie, A.; Seidel, J.; Rijnbout, S.; Waheed, A.; Schmidt, B.; Geuze, H.; von  
J. Biol. Chem. 264, 13865-13872, 1989  
A;Title: Cloning and expression of human steroid-sulfatase. Membrane topology, glycosylation  
A;Reference number: A32641; MUID:89340479  
A;Accession: A32641  
A;Molecule type: mRNA  
A;Residues: 1-583 <STE>  
A;Cross-references: GB:J04964; NID:g338564; PIDN:AAA60597.1; PID:g338565  
A;Experimental source: BHK-21 cells  
A;Note: parts of this sequence were determined by protein sequencing  
R;Yen, P.H.  
unpublished results 1988, cited by GenBank  
A;Reference number: A57116  
A;Accession: A57116  
A;Molecule type: mRNA  
A;Residues: 1-22, 'E', 24-583 <VEN1>  
A;Cross-references: GB:M16505; NID:g338513; PIDN:AAA60596.1; PID:g338514  
R;Yen, P.H.; Allen, E.; Marsh, B.; Mohandas, T.; Wang, N.; Taggart, R.T.; Shapiro, L.J.  
Cell 49, 443-454, 1987  
A;Title: Cloning and expression of steroid sulfatase cDNA and the frequent occurrence of  
A;Reference number: A25961; MUID:87187642  
A;Accession: A25961  
A;Molecule type: mRNA  
A;Residues: 1-22, 'E', 24-456, 'LRTTHSPGRFPSPPTSTPWFQRIALPHTCASVSGVMSPTTHLYSLIPP', 'KIPERR  
A;Cross-references: GB:M16505; NID:g338513  
A;Note: this sequence revised in A57116  
A;Note: part of this sequence, including the amino end of the mature protein, was determined  
R;Kawano, J.I.; Kotani, T.; Ohtaki, S.; Minamino, N.; Matsuo, H.; Oinuma, T.; Aikawa, E.  
Biochim. Biophys. Acta 997, 199-205, 1989  
A;Title: Characterization of rat and human steroid sulfatases.  
A;Reference number: S05414; MUID:89352671  
A;Accession: S05415  
A;Molecule type: protein  
A;Residues: 22-43, 'X', 45-46 <KAW>  
R;Dibbelt, L.; Otto, J.; Kuss, E.  
Biol. Chem. Hoppe-Seyler 370, 847-848, 1989  
A;Title: The N-terminal amino-acid sequence of human placental sterol sulfatase.  
A;Reference number: S05423; MUID:90074181  
A;Accession: S05423  
A;Molecule type: protein  
A;Residues: 22-35 <DIB>  
R;Yen, P.H.; Marsh, B.; Allen, E.; Tsai, S.P.; Ellison, J.; Connolly, L.; Neiswanger, K.  
Cell 55, 1123-1135, 1988  
A;Title: The human x-linked steroid sulfatase gene and a Y-encoded pseudogene: evidence  
A;Reference number: I52800; MUID:89077541  
A;Accession: I52800  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 134-274 <YEN3>  
A;Cross-references: GB:M23945; NID:g338604; PIDN:AAA60598.1; PID:g338607  
A;Accession: I65619  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 461-583 <YEN4>  
A;Cross-references: GB:M23556; NID:g338605; PIDN:AAA60599.1; PID:g338608  
C;Genetics:  
A;Gene: GDB:STS  
A;Cross-references: GDB:120393; OMIM:308100  
A;Map position: Xp22.32-Xp22.32  
A;Note: Defects in this gene can cause X-linked ichthyosis  
C;Function:  
A;Description: hydrolyzes 3beta-hydroxysteroid sulfates to release sulfate  
C;Superfamily: animal sulfatase  
C;Keywords: endoplasmic reticulum; glycoprotein; lysosome; microsome; sulfonic ester hydrolase  
F;1-21/Domain: signal sequence #status predicted <SIG>  
F;22-583/Product: steryl-sulfatase #status predicted <MAV>  
F;185-211/Domain: transmembrane #status predicted <TMI>  
F;213-237/Domain: transmembrane #status predicted <TM2>  
F;47, 259/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F;75/Modified site: 3-oxoalanine (Cys) #status predicted  
F;333,459/Binding site: carboxylate (Asn) (covalent) #status absent

Query Match 13.7%; Score 413; DB 1; Length 583;  
Best Local Similarity 25.1%; Pred. No. 1-2e-23;  
Matches 142; Conservative 78; Mismatches 201; Indels 144; Gaps 22;  
QY 67 EPSTTTSQPHILIPILADDOGFRDVGHGSE-IKTPDLKLAAGVKL-ENYVQPICTP 124  
Db 18 EASHAASRPNILVMADDLIGDPCGYNKRTIRTNIDRLASGGVKLTOLHAAASPLCTP 77  
QY 125 SRQFINGKVIHTGLOH-----SIIRTPQNCPLDNLATLPQKLKEVGSTHVMGKWHL 179  
Db 78 SRAAFMTGRYPVRSGMASWSRTGVFLFTASSGGGLPTDEITFAKLLKDGQSTALIGKWHL 137  
QY 180 GFYRKEC-----MPTRRGDFTFGSL-----GSGDYIT----- 208  
Db 138 GM---SCHSKTDFCHHPLHGFNYFYGISLTNLROCKPBGSVFTTGKRLVFLPLQIVG 194  
QY 209 -----HY-----KCDSPMCGYDLYE 224  
Db 195 VTLLTAAALNCLGLLHVPLGVFFSLLELAALILTLFLGLHYFRPLNCFM--MRNYEIIQ 252  
QY 225 NDNAWDYDNGIYSTOMYTORVQOILASHNPTKPIFLYIAVQAVHSPLOAPGRYFEHYS 284  
Db 253 Q---PMSYDN---LTQRLTVEAAQFI-ORNTETPFLVLVSLVHLVHTALFSSKDFAGKSQH 305  
QY 285 IININRRRYAAMLSCLDEANNVTALKTGYFYNNSTIIYSSDNGGQ-----PTAGG 336  
Db 306 GV-----YGDVEEDMSVQIILNLLDELRLANDTLIYFTSDOGAHVEEVSSKGEIHGG 359  
QY 337 SNMPLRSGKTYEGGIRAVGVFVHSPLLKNGTVCKPEVHITDWTYPTLISLAEGQIDEDI 396  
Db 360 SNGIYKGGKANWEGGIRVPGILRWPRVIOAQKIDEPSTNMDFPTVAKLAGAPLPEDR 419  
QY 397 QLDGYDIWETISEGLRSPRD--ILHNIDPIYTKAKNGSWAAGYGIWNTAQSAIRVQHW 454  
Db 420 IIDGRDLMLPL-EG-KSORSDFELFYHCNAYLNVR-----WHPQNSTSI-----W 464  
QY 455 KLLTGNPGYSDWVPQPSNLGNPRWNERITTSSTGKSV-----WLFNITADPVERVDL 508  
Db 465 KAFFETP-----NFPNVGSGCFATHVCFCGVSVTHHDPPLLFIDISKDPRERNPL 515  
QY 509 S-----NRYPGIVKKLRLRLSQFNKT 529  
Db 516 TPASEPRFYEILKVMQEAADRHTQT 540  
RESULT 10  
S69336  
arylsulfatase (EC 3.1.6.1) - Pseudomonas aeruginosa  
C;Species: Pseudomonas aeruginosa  
C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Mar-2000  
C;Accession: S69336; S69359  
R;Beil, S.; Kehrl, H.; James, P.; Staudenmann, W.; Cook, A.M.; Leisinger, T.; Kertes  
Eur. J. Biochem. 229, 385-394, 1995  
A;Title: Purification and characterization of the arylsulfatase synthesized by Pseudo  
A;Reference number: S69336; MUID:95262702  
A;Accession: S69336  
A;Molecule type: DNA  
A;Residues: 1-533 <BEI>  
A;Cross-references: EMBL:248540; NID:g2440146; PIDN:CAA88421.1; PID:g695684  
A;Experimental source: strain PAOI  
A;Accession: S69359  
A;Molecule type: protein  
A;Residues: 2-27,95-101,131-155;159-169;188-195;245-263;297-302;340-350;383-407-4  
R;Dierks, T.; Misch, C.; Hummerjohann, J.; Schmidt, B.; Kertesz, M.A.; von Figura, K.  
J. Biol. Chem. 273, 25560-25564, 1998  
A;Title: Posttranslational formation of formylglycine in prokaryotic sulfatases by mo  
A;Reference number: A59074; MUID:98421466  
A;Contents: annotation; post-translational modification  
A;Note: confirmation of 3-oxoalanine active site, referred to as formylglycine  
C;Genetics:  
A;Gene: atsa  
C;Superfamily: animal sulfatase

A: Cross-residues: 1-389 <RES>  
A: Cross-references: EMBL:X83573; NID:g791003; PIDN:CAA58556.1; PID:g791004  
C:Genetics:  
A:Gene: GDB:ARSE; CDPX; CDPX1  
A:Cross-references: GDB:555743; OMIM:302950  
A:Map position: Xp22.3-Xp22.3  
C:Superfamily: animal sulfatase  
C:Keywords: sulfuric ester hydrolase  
F:86/Modified site: 3-oxoalanine (Cys) #status predicted

Query Match	13.2%	Score 399;	DB 2;	Length 593;
Best Local Similarity	24.4%	Pred. No. 1.5e-22;		
Matches 144;	Conservative	87;	Mismatches 191;	
			Indels 168;	Gaps 27;

Qy	67	EPSTTSTSQPHLFIILADDOGFDRDVGHGSE-IKTPTLTKLAEGVKL-ENYYVQICPTP	124
Db	32	EPTANAFKNILLINADLGTGDLGCGYGNNTLRPTNIDQABEGVRLTQHLAAAPLCTP	91
Qy	125	SRSQFIITGKYOIHTGLQHSI-IRTPQNC-----LPLDNATLPKLEKVCYSTHMGCKWHL	179
Db	92	SRAAFLTGRHSFRSGMDASNGYRALQWNAAGSGGLPENETTFARILQOHHGYATGLICKWHQ	151
Qy	180	GFYRKECM-----PTRRGFTDFGSGLLSGSDYTHVKCDSPGCMGYDLVDNNA---	228
Db	152	GV---NCASRGDCHHPLNHGFDYFCM-----PFLITNDCD-PGRPP---EVDAAALA	198
Qy	229	-ANDYN---GI-----YSTMOTQRVOQL	250
Db	199	QLMGYQTFALGILTLAAGQTCGFVSARAVTGMAGVGLCFITSWSGFVRRNCIL	258
Qy	251	-ASHNPTK-----PFLYIATQAVHSPILQACGRYF---E	280
Db	259	MRNHDTVEQPMVLKLEKTASLMKKAVSYIERKHKGHPFLFLSLHLVHIPLVTTSAFLGKSQ	318
Qy	281	HYRSIINRRRYAAMLSCLDENANNVTALKTYGYFNNSIIYISDONGQOPTA-----	334
Db	319	H-----GLGDNVEEMDWLIGKVLNAIEDNGLKNSTFTFTSDHGGHLEARDCHSQ	369
Qy	335	-GGSNWP LRSGKGT-YWEGIRAVGFVHSPLLKNKGTVCKEPVHIITDWTPLISLAEGOI	392
Db	370	LGGWNGYIKGKGKMGWEGGIRVPGIFHWPGLVPAGRVIGEPTSLMDVPTTVOLVGGEV	429
Qy	393	DEDIQLDGYDIWETISBG-LRSRPVDILHNIDPIYTKAKNGSWAAGYGTWNTAQSAIRV	451
Db	430	PQDRVIDGHSVLPLLOGAEARSAAHEFLFYHCGQHLHAAR-----WHQKDSGSV---	477
Qy	452	QHWKLLTGNPCYSDWVPQSFNSLGNPNRHNER-----ITSSTCKSYW-----LFN	497
Db	478	--HKV-----HYTTPQ-----FPIPERGILLTAEASAHAEWGVTTHRPPFLFD	518
Qy	498	ITADPYE-----RVDLSNRPYPGIVKLLRRLSQFNKTAVPVRYPPKDPRSN	543
Db	519	LSRDPSSARPLTDPDEPLYHAVIARYGAAYSEHROTLSPV---POOF-SMNS	566

RESULT 13  
C91220  
arylsulfatase [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C:Accession: C91220  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: C91220  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-551 <HAY>  
A:Cross-references: GB:BA000007; PID:BA038154.1; PID:gl3364207; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Genetics:  
A:Gene: ECs4731

```

Query Match      13.2%; Score 398; DB 2; Length 551;
Best Local Similarity 24.6%; Pred. No. 1.6e-22;
Matches 148; Conservative 97; Mismatches 204; Indels 152; Gaps 29;

Qy      2  APRCAGHPP-----PSFQACVCGKMLAMGALAGFWILCLLTGYLSWGQALBEEEGE 57
      | : | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      33  ARKFAGYDHPNQYLKVPATTIADNMMPVMQHPA-----QDKPEQQ 73

Qy      58  LLAQAGEKLEPSTSTSQPHLIFLADDOGRDVGYHGSEI-----KPTFLDKLAEGVKL 113
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      74  KLALEKK-----TGKKPNVVFLLDDVGWMDVFGNGGVAVGNPTPDIDAVASOGLI 127

```

**Qy** 114 ENYVQPICTSRSQFITGKYIHTGLQHSIRPT---QPNCPLDNLATLPQKLKEYGS 170  
:  
:  
:  
**Db** 128 TSAYSQSPSSPTRATILTGQYSIH---HGILMPMYGPQGLO-GLTTLPOLLHDCGYV 182  
:  
:  
:  
**Qy** 171 THMWKGWHLGYRECMPTRRGFDTF--FGSLGSGDYHYTHKCDSPGMCGCYDIYENDNA 228  
:  
:  
:  
**Db** 183 TOAIGKHMG-ENKESQPQNYGDFDFRGNSV---SDMYTEWR-----DVHVNPVEV 229  
:  
:  
:  
**Qy** 229 AWDVDNGIYSTQMVTOR-----VOOILASHNP----- 255  
:  
:  
:  
**Db** 230 ALSPDRSEYIKQLPSKDDVHAVRGGEOQAADITPKYMEDLDQRWMEYGVKFLDKMAKS 289  
:  
:  
:  
**Qy** 256 TKPIFLXIYAQVHSPLOAGPYFEHYRSI----ININRRRYAAMLSCLEAINNVTLAL 311  
:  
:  
:  
**Db** 290 DKPFPLYGTGRGH-----FDNYPNAYAGSPPARTSYGDMVMEMDNVFANLYKAL 340  
:  
:  
:  
**Qy** 312 KTYGFYNNSIIYSSDNGQ---PTAGSNPNRLSKGTWYEGGIRAVGFVHSPLLKNKG 368  
:  
:  
:  
**Db** 341 EKNQOLDNTLIVTSDNGPEAEVPHPGRT--PFRGAKGSTWEGGVVRPTFVVW-----KG 393  
:  
:  
:  
**Qy** 369 TVCKEP-----VHTDWYPTLLISLAEQIDEDIQLDQYDIWETISEGLRSPRVDLHNI 422  
:  
:  
:  
**Db** 394 MI--QPRKSDIGVDLADLFPTALDLA-GH-----PGAQVANLV-----PKTFIDGV 437  
:  
:  
:  
**Qy** 423 DP-----ITYKKAGNSWAAGYGIWNATAQSATRVQHVKLLTCGNPGYSMDV--PPQSFSLGP 477  
:  
:  
:  
**Db** 438 DOTSFFLGTNGQSRKAHEYFL--NGKLAAYRMDEEK-----YHVLIQQPAYATOSGY 488  
:  
:  
:  
**Qy** 478 NRWHNERITSSTGKSVWLFNITADPYRVDLSNRYPGIVKLLRRLSQFNKTAPVPVRYPP 537  
:  
:  
:  
**Db** 489 QGGTGTVMQTAGSSV--FNLYTDQESDSGVRHPHMGVPLQTEMHAYME--ILKKYPP 544  
:  
:  
:  
**Qy** 538 K 538  
:  
**Db** 545 R 545

**RESULT 14**

E86066  
arylsulfatase [imported] - Escherichia coli (strain O157:H7, substrain EDL933)  
C:Species: Escherichia coli  
C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: E86066  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May  
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: E86066  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-551 <STO>  
A:Cross-references: GB:AF005174; NID:g12518665; PIDN:AAG58993.1; GSPDB:GN00145; UWGP:  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
C:Gene: asia

Query Match	13.2%	Score 398;	DB 2;	Length 551;
Best Local Similarity	24.6%;	Pred. No. 1.6e-22;		
Matches 148;	Conservative	97;	Mismatches 204;	Indels 152;
Gaps 29;				

[illegible]

Db 128 TSATSQSPSSPTRATIITGQSYIH----HGIILPPMTQGPGGLQ-GLTTLPLLHDGGYV 182

Qy 171 THMVKGWHLGYRKECKPRTGRGDTF--FGSLIGSGDYTYHYKCDSPGMCGYDLIENDNA 228  
| : | | | | : | | | | : | | | : | | : | : |  
Db 183 TQAICKHMG-ENKESQPQNVGFDFRGNSV---SDMYTEWR-----DVHVNPV 229

Qy 229 AWYDNGNIYSTOMYTQR-----VOQILLASHNP----- 255  
| : | | : | : | | | | | : | | : | | :

Db 230 ALSPDREXYIKQLPFKSDDVHAVRGEGQAIAIDTPKYMEDLDQRMWEYGVKFLDKMAKS 289

Qy 256 TKPIFLXIYAQVHSPLQAPCRFEHYRSI-----ININRRRYAAMLSCLEAINNVTAL 311  
| : | | | | : | : | | | | | : | | : | : | : | : | |

Db 290 DKPEFLTGTGRGCH-----FDNPNAYAGSSPARTSYGDCVMEMDNVFANLYKAL 340

Qy 312 KTYGFYNNSIIISDNGGO---PTAGGSNWPLRGSTGWEGIRAVGVFVSPLLNKG 368  
| : | : | : | : | | | | | : | | | | : | | | : | | : | |

Db 341 EKNQOLDNTLIVFTSNDNGPEAEVPHGRT--PFRGAKGSTWEGVRVPYFW-----KG 393

Qy 369 TVCKEP-----WHITDWYPTLSLAEQIDEDIQLDGVDIWETISEGLSRPRVIDLIHI 422  
| : | | : | : | : | | | | : | | : | | : | : | : |

Db 394 MI--OPRKSDGIVDLADLFPTALDLA-GH-----PGAUVANLV-----PKTTFDGV 437

Qy 423 DP-----IYTKAKGSWAAGVGIIWNATQSARVQHVKLLTGNGPYSDWV--PPQSFNLGP 477  
| : | : | : | : | | | | | : | | : | : | : | : | : |

Db 438 DOTSFFLGTSNQSNKRKAHEHYFL--NGRLAAVRMDDEF-----YHVLIOOPYAYTOSG 488

Qy 478 NRWHNERITSTGSKSVLENLTADPYERVDLSNRYPGVIGVKLLRRLSQFNKTAPVPYRP 537  
| : | : | | : | | | : | | | : | | : | : | : | : | : |

Db 489 QGGFTGVMTAGSSV--FNLYTDQSDSIGVRHPWPVLPQTEMHAYWB--ILKKYPP 544

Qy 538 K 538  
:

Db 545 R 545

RESULT 15

F95993

C:Probable arylsulfatase (EC 3.1.6.1) [imported] - Sinorhizobium meliloti (strain 102), m

C:Species: Sinorhizobium meliloti

C>Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 14-Sep-2001

C:Accession: F95993

R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan

Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A:Title: The complete sequence of the 1.683-kb pSymB megaplasmid from the N2-fixing endo

A:Reference number: A95842; MUID:21396508; PMID:11481431

A:Accession: F95993

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-537 <CUR>

A:Cross-references: GB:AL591985; PID:NACAC49614.1; PID:g15141101; GSPDB:GN00167

A:Experimental source: strain 1021, megaplasmid pSymB

R:Galibert, F.; Finan, T.M.; Long, S.R.; Puibler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,

pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.

L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,

hebaullt, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.

A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation

C:Genetics:

A:Gene: aslA; SMB20915

A:Genome: plasmid

C:Keywords: sulfuric ester hydrolase

Qy	120	PICTPSRSQFTTGKYQIHTGLQHSIIIRPTOPNCLPLDNATLPKQKKEVGYSTHVMGKWHL	179
Db	103	QSCTAGRAAFTGSGSPRTGLT-KVGLPGADIGIQPEDATVAELLKSLGYATQFGKNHL	161
Qy	180	GFYKCEMPTRRRFDTPFGSLGSDYYTHYKCDSPGMCYDLYENDNAAMDYDNGIYST	239
Db	162	G-DKDEFLLPTAHGDEFEGNLY-----HUNAE-----EPENPDYPODPA-FRK	203
Qy	240	QMYTORVOQIILASH--NPTKPI-----FLYTAYQAVHSPLOAQPRGYEFHYRSI-	285
Db	204	RFGPRGVIKATADCKIEDTGPLSVKRMETVDQEFLEALDFIDRKAKGDAPWFCYFNSTR	263
Qy	286	--ININRRRYAAMLSC-----DEAINNVTLALKTYGYFNNSIIYSSDNGQOPT	333
Db	264	MHVANTHLKPDPSAGTKGLGVYPDGMVEHDGHVGLKLLKLDLGLTENTIVVYTSDNCAEVM	323
Qy	334	A--GGSNWPLRSGSKGTWYEGGIRAVGFVHSPPLKKNKTCVCKEPVHTDMWPT-----	383
Db	324	TWPDGGNTPFGRKATNWECSGFRVPMCIWRPGVLIKPGCIIHNERPFSHYDLIPTCAAAGEP	383
Qy	384	-----LISLAEGQIDEDIQLDGYDIWETISEGLR-SPRVDIILHNTDPIYTKAKNGSWAA	436
Db	384	DIVAKCLTGYAAGAKTFKVHLGDYGNLMPFLSGSSNDAPRRDFLY-----	427
Qy	437	GYGTWNTAIQ-SATRVQHWKLL-----TCNPGYSVDWVPQSPFNLGNPNRWHNERITSSTGKS	492
Db	428	---WDDGELVAVRVQNNKVVFKFSQDHGIGVWRQP-FTEL-----RA	466
Qy	493	VWLFNITADPYERVDLSNRY-----PGIVKLLLRLLSQFNKTAVPVRYPP	537
Db	467	PLLFNLRSDPERGDTMEYKEWFFDFRSFVVVPSQAAVVAKWLSEFKF-----PIROKP	520

Search completed: July 23, 2002, 19:04:13  
Job time: 6549 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 15:04:54 ; Search time 4089.32 Seconds

(without alignments)  
11595.946 Million cell updates/sec

Title: US-09-495-823-8

Perfect score: 2266

Sequence: 1 cagcgtccgcacacgcgtc.....tgccacctggtgcgaattc 2266

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_hgt.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vl.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vl.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
------------	-------	-------	--------	----	----	-------------

RESULT 1

AX206967

LOCUS

AX206967

DEFINITION

Sequence 8 from Patent WO0155411.

AX206967

ACCESSION

AX206967.1

VERSION

GI:15394724

KEYWORDS

human.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 2253)

AUTHORS

Glucksmann, M.A., Williamson, M., Rudolph-Owen, L.A. and Tsai, F.Y.

TITLE

Human sulfatases

JOURNAL

Patent: WO 0155411-A 8 02-AUG-2001;

Millennium Pharmaceuticals, Inc. (US)

FEATURES

Location/Qualifiers

1..2253

source

/organism="Homo sapiens"

/db\_xref="taxon:9606"

324..2123

CDS

/note="unnamed protein product"

/codon\_start=1

ALIGNMENTS

AX206967

Sequence 8 from Patent WO0155411.

AX206967

AX206967.1

GI:15394724

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 2253)

AUTHORS

Glucksmann, M.A., Williamson, M., Rudolph-Owen, L.A. and Tsai, F.Y.

TITLE

Human sulfatases

JOURNAL

Patent: WO 0155411-A 8 02-AUG-2001;

Millennium Pharmaceuticals, Inc. (US)

FEATURES

Location/Qualifiers

1..2253

source

/organism="Homo sapiens"

/db\_xref="taxon:9606"

324..2123

CDS

/note="unnamed protein product"

/codon\_start=1



Db 1861 CAGGAATCGTGAAGAAGCTCCTACGGAGGCTTCACAGTTCAACAAAACTGCAGTGC GG 1920  
Qy 1921 Caggtatcccccaagaccaccccaagaaagtaaaccttagctcaatgaggggtctcaggac 1980  
Db 1921 TCAGGTATCCCCCCANAGACCCCAAGAACTAACCTTAGCTCAATGAGGGGTCTGGGAC 1980  
Qy 1981 catggtatagagagaaaccacgaagaaagacaaagcaaaatcaggctgagaaaaagc 2040  
Db 1981 CATGTTATAAGAGAGAAACCAAGAAAAGAACGCAAGCAAAAATCAGCTCAGAAAAAGC 2040  
Qy 2041 aaagaaaagc-aaaaaaagaaagaaagaaagcagaaagcaggtctcaggttcaactggc 2099  
Db 2041 AAAAGAAAAGCAAAAAAAGAAAGAAACAGCAGAAAAGCAGTCTCAGGTTCAACTGCC 2100  
Qy 2100 attcaggtgtactgtgtgataagcacaaatatcttctgttgggttaaacctttaatcagt 2159  
Db 2101 ATTCAAGTGTACTGTGTGATAGACAAATATTTCTGTGTTAAACATTTAATCAGT 2160  
Qy 2160 tcttatcttcatctgtttcttagtaaacccagcaaaatttgctcgataaatcgtggc 2219  
Db 2161 TCTTATCTTTTCATCTGTTTCTTAGTAAACCCAGCAAAATTTGCTCGATAATATCGTGGC 2220  
Qy 2220 ctacgctcagggctgtgttttcatcgtctgcccac 2252  
Db 2221 CTAAGCGTCAGGCTGTGTTTTCATGCTGTGCCAC 2253

RESULT 2  
AX206973 AX206973 1800 bp DNA linear PAT 30-AUG-2001  
LOCUS Sequence 14 from Patent WO0155411.  
DEFINITION AX206973  
ACCESSION AX206973  
VERSION AX206973.1 GI:15394729  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1800)  
AUTHORS Glucksmann,M.A., Williamson,M., Rudolph-Owen,L.A. and Tsai,F.Y.  
TITLE Human sulfatases  
JOURNAL Patent: WO 0155411-A 14 02-AUG-2001;  
Millennium Pharmaceuticals, Inc. (US)  
FEATURES  
source  
1..1800  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 534 a 434 c 423 g 409 t  
ORIGIN

Query Match 78.6%; Score 1781.6; DB 6; Length 1800;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1795; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

Qy 324 atggctccacagggctgtgcggggcatccgctccgcttctccacagggctgtgtctgt 383  
Db 1 ATGGCTCCACAGGGGTGTGCGGGGATCCGCCCTCGCCCTTCCACAGGCCGTGTCTGT 60  
Qy 384 cctgaaagatgctagcaatggggcgctgcgaggaattcgtggtcctgctcact 443  
Db 61 COTGAAAGATGCTAGCAATGGGGCGGTGCGAGGATTCGTGATCCTCTGCTCCTCACT 120  
Qy 444 tatgtttacctgtctctgggcagcccttagaagaggaggaagggccttactagct 503  
Db 121 TATGTTTACCTGTCTCTGGGCGAGGCCCTTAGAAGAGGAGGAAGAGGGGCCCTTACTAGCT 180  
Qy 504 caagctggagagaaactagagccagcacaaacttccactcccgcccaactctatttc 563  
Db 181 CAAGCTGGAGAGAACTAGAGCCCGACACACTTCCACCTCCCGCCCATCTCATTTTC 240  
Qy 564 atcctagcggatgatacagggatttagagatgtgggttaccacggatctgagattaaaaa 623

Db 241 ATCTAGCGGATGATCAGGGATTTAGAGATGTGGGTTACCACGGATCTGAGATTAAACA 300  
Qy 624 cctactcttgaagctcgctgcgaagagacttaaaccttggaagaaactactatgtccagcct 683  
Db 301 CCTACTCTTGACAAGCTCGCTGCGGAGAGGTAAACTTGGAGAACTACTATGTCTCAGCCT 360  
Qy 684 attgcacaccatccaggagtcagtttatctactggaagtatcagatcacacacggactt 743  
Db 361 ATTTGCACACCATCCAGGAGTCAGTTTATTACTGGAAGTATCAGATACACACGGACTT 420  
Qy 744 caacattctataagacctaccaccccccaactgtttacctctggacaaatgcaccccta 803  
Db 421 CAACATTTCTATATAAGACCTACCCCAACCCCAACTGTTTTACCTCTGGCAATATGCCACCTA 480  
Qy 804 cctcagaacactgaaggaggttgatatattcaacgcataatgctcgaaaaatgcacttgggt 863  
Db 481 CCTCAGAAACTGAAGGAGGTGGATATTCAACGCATATGGTCGGAATGGCATTGGGT 540  
Qy 864 ttltacagaaagaatgcatgcccaccagaagagatttgatacccttttttgggtccctt 923  
Db 541 TTTTACAGAAAGAAATGCATGCCACACAGAGAGGATTTGTATACCTTTTGGTTCCCTT 600  
Qy 924 ttgggaagtggggttactatacacactacaaaatgtgcagctcctcgggagtgtggctat 983  
Db 601 TTGGGAAGTGGGATTTACTATACACACTACAAATGTGACAGTCTCTGGGATGTGGCTAT 660  
Qy 984 gacttgatgaaacgaatgctgctgggactatagaacatagaacatacactccacacag 1043  
Db 661 GACTGTGTATGAAAACGACAATGCTGCCCTGGGACTATGACAAATGGCATATACTTCCACAG 720  
Qy 1044 atglacactcagagtagtacagcaaatcttagcttcccataacccccacaaagcctatttt 1103  
Db 721 ATGTACACTCAGAGAGTACAGCAAAATCTAGCTTCCCAATAACCCCAAGAGCTATATTT 780  
Qy 1104 ttatatgtcctatcaagctgttcaatccacactgcaagctcctcctcggcaggtatttcgaa 1163  
Db 781 TTATATATTTGGCTATCAAGCTGTTTCATTCACCACCTGCAAGCTCTCTGGCAGGTATTTGCAA 840  
Qy 1164 cactaccgactcattatacaataaacagagagagatatgctgcacatcttctcgtcta 1223  
Db 841 CACTACCATTCATTATCAACATAAACAGGAGGAGATATGTCGCATGCTTTCCTGCTTA 900  
Qy 1224 gatgaagcaatacaacacgtgacattggctctaaagacttatggtttctatacaaacagc 1283  
Db 901 GATGAAGCAATCAACAACGTCGACATTGGCTCTTAAAGACTTATGTTTCTATAACAACAGC 960  
Qy 1284 attatcattactcttcagataaatgtggccagcctacgagggagggagtaactggcct 1343  
Db 961 ATTATCATTTACTCTTCAGATAAATGGTGGCGAGCCCTACGGCAGGAGGAGTAACCTGGCCT 1020  
Qy 1344 ctacaggtagcaaaagaaacatatattgggaagaggatccgggctgtaggctttgtgcat 1403  
Db 1021 CTCAGAGGTAGCAAAAGAAACATATTGGGAAGAGGGATCCGGGCTGTAGGCTTTGTGTCAT 1080  
Qy 1404 agccactcttgaaaaacaaggaacagctgtgtaaggaaacctgtgcacatcaactgactgg 1463  
Db 1081 AGCCCACTTCTGAAAAACAAGGGAACAGTGTGTAAGGAACCTGTGCACATCACTGACTGG 1140  
Qy 1464 taccacctctcatttcaactggctggaagacagatgtgatgggacattcaactagatggc 1523  
Db 1141 TACCCCACTCTCATTTTCATCTGGCTGAAGACAGATTTGATGAGGACATTTCAACTAGATGGC 1200  
Qy 1524 tatatatctggagaccataagttagggctcttcgctcacccccctagatattttgcat 1583  
Db 1201 TATGATATCTGGAGACCATTAGTGAGGGGTCTTCCTCACCCCGAGTAGATATTTTGGAT 1260  
Qy 1584 aacattgaccccatatacaccaagcaaaaaatggctcctcggcgagcagggctatgggac 1643  
Db 1261 AACATTGACCCCATATACACCAAGCAAAAAATGCTCCTCTGGGCAGCAGGCTATGGGATC 1320  
Qy 1644 tggaaacactgcaatccagtcagccatcagagtgcagcactggaaattgcttacaggaat 1703

Db 1321 TGAACACTGCAATCAGCTACGCCATCAGAGTCAGCACTGGAAATTTGCTTACAGAAAT 1380  
QY 1704 cctgctacagcagctgggtccccctcagctctttcagcaaacctggaccgaacggtagg 1763  
Db 1381 CCTGGCTACAGCGACTGGGTCCCCCTCAGCTTTTCAGCAACCTGGGACCGGAACGGTGG 1440  
QY 1764 cacaatgaacgagatcacctcgtcaactggcgaagtgtatgcttttcaacatcacagcc 1823  
Db 1441 CACAATGAACGGATCACCTTGTCAACTGGCAAAAGTGTATGGCTTTTCAACATCACAGCC 1500  
QY 1824 gaccatatgaaggggtggacctatctaacagggtatccagggaatcgtgaagaagtctcta 1883  
Db 1501 GACCCATATGAGAGGGTGGACCTATCTAACAGGTATCCAGGAATCGTGAAGAAGTCCTA 1560  
QY 1884 cgaagcctctcacagtccaacaaaactcagtcgagccggtcaggtatcccccaagacccc 1943  
Db 1561 CGAGGCTCTCAGAGTTCAACAAAACACTCGAGTGGCGGTCAGGTATCCCCCAAGACCCC 1620  
QY 1944 agaagtaacccttagctcaatggaggggtcttagggaccatggtatagagaggaacccaag 2003  
Db 1621 ACAAGTAACCTTAGCTCAATGGAGGGGTCTGGGACCATGTGTATAAGAGGAAACCAAG 1680  
QY 2004 aaaaagaagcgaagcaaaaactcagctcagaaaaagcaaaaagaagc-aaaaaagaag 2062  
Db 1681 AAAAAAGAGCCCAAGCAAAAATCAGGCTCAGAAAAAGCAAAAAGCAAAAAGAAAG 1740  
QY 2063 aagaacagcagaagcagctcaggttcaacttccattcaggttacttctgtgataa 2122  
Db 1741 AAGAAACAGCAGAAAGCAGCTCAGGTCAACTTCCATTACAGGTGTACTTGTGGATAA 1800  
  
RESULT 3  
AC104779/c  
LOCUS Homo sapiens chromosome 4 clone RP11-26P13, linear PRI 09-FEB-2002  
DEFINITION Homo sapiens  
ACCESSION AC104779  
VERSION AC104779.4 GI:18642899  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 152861)  
Waterston,R.H.  
The sequence of Homo sapiens clone  
2 (bases 1 to 152861)  
Waterston,R.H.  
Direct Submission  
Submitted (21-DEC-2001) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
3 (bases 1 to 152861)  
Waterston,R.H.  
Direct Submission  
Submitted (09-FEB-2002) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On Feb 9, 2002 this sequence version replaced gi:18482303.  
  
----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site:http://genome.wustl.edu/gsc/index.shtml  
Contact: submissions@watson.wustl.edu  
----- Project Information -----  
Center project name: H\_NH0026P13  
----- Location/Qualifiers -----  
1. 152861  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="4"  
  
FEATURES  
source

BASE COUNT 48870 a 27681 c 28505 g 47805 t  
ORIGIN  
  
Query Match 66.9%; Score 1515; DB 9; Length 152861;  
Best Local Similarity 99.6%; Pred. No. 0;  
Matches 1529; Conservative 0; Mismatches 5; Indels 1; Gaps 1;  
  
QY 719 aaagtatcagatacacacacgggacttcaacattctatcatgaagacctaccaccaactg 778  
Db 10199 AAGGTATCAGATACACACCGGACTTCAACATTTCTATCAAGACCTTACCAACCAACTG 10140  
QY 779 ttactctctggacaatgccaccctacctcagaaactgaagaggttggtatattcaacgca 838  
Db 10139 TTTACTCTGGACAATGCCACCCTACCTCAGAAACTGAAGGAGTTGGATATTCAACGCA 10080  
QY 839 tatgctcggaaaatggcacttgggtttttacagaaaagaatgcattgcccaccagaagag 898  
Db 10079 TATGCTCGAAAAATGGCACTTGGGTTTTTACAGAAAAAAGTAATGCATGCCACCAGAAGAGG 10020  
QY 899 atttgatacccttttttgggtcccttttgggaagtggggaattactatacacactacaaatg 958  
Db 10019 ATTTGATACCTTTTTTGGTTCCCTTTTGGGAAAGTGGGATTACTATACACTACAAATG 9960  
QY 959 tgacagctcctgggatgtgtgctatgacttgcataaaacgacaatgctgctggacta 1018  
Db 9959 TGACAGTCTCGGGATGTGGGCTATGACTTGTATGAAACGACAATGCTGCCTGGACTA 9900  
QY 1019 tgacaatggcatatactccacacagatgtacactcagagagtacagcaaatcttagcttc 1078  
Db 9899 TGACAAATGGCATATACTCCACACAGATGTACACTCAGAGAGTACAGCAATCTTAGCTTC 9840  
QY 1079 ccataaccccccaagcctatatattttatatatttgcctatacgaagtgttcattccaccat 1138  
Db 9839 CCATAACCCCCACAAAGCCTATATTTTATATATTTGCTATCAAGCTGTTTCATTCCACT 9780  
QY 1139 gcaagctcctggcaggtatttcgaacactaccgattccattatacaataaacaagagag 1198  
Db 9779 GCAAGCTCTGGCAGGATTTTCGAACACTTACCAGTCCATTATCAACNTAAGCAGGAGG 9720  
QY 1199 atatgctgccatgcttctcgtcttagaagaagaatcaacaacgtgacattggtcctaaa 1258  
Db 9719 ATATGCTGCCATGCTTTCTCTGCTTAGATGAAGCAATCAACAACGTGACATTGGCTCTAA 9660  
QY 1259 gacttatggtttctataacaacacagcattatacttactctcagataaatggtgcagcc 1318  
Db 9659 GACTTATGCTTTCTATAACAACAGCATTTATCATTTTACTCTTCAGATAAATGGTGGCCAGCC 9600  
QY 1319 tacgacagagagagtaactcgtcctcagaggtagcaaggaacacatatgggaagagag 1378  
Db 9599 TACGGCAGGAGGAGTAAGTGGCTTCTCAGAGGTAGCAAGGAAGAACATATTTGGGAAGGAGG 9540  
QY 1379 gatccgggctgtaggcttttggcatagcccacttctgaaaaacaagggaacagtgtgtaa 1438  
Db 9539 GATCCGGGCTGTAGGCTTTGTGCATAGCCCACTTCTGAAAAACAAGGGAACAGTGTGTA 9480  
QY 1439 ggaacctgtgacatcaactcagctggtgtaccccaactctcatttcactggtcgaaggacagat 1498  
Db 9479 GGAACCTTGTGCACATCACTGACTGGTGTACCCCACTCTCATTTTTCACCTGGCTGAAGGACAGAT 9420  
QY 1499 tgatgagacattcaactagatggctatgatctgggagaccataaagtggaggtcttcg 1558  
Db 9419 TGATGAGGACATTCAACTAGATGGGTATGATATCTGGGAGACCAATAAGTAGGGTCTTCG 9360  
QY 1559 ctcaaccccgagtagatattttgtcataacattgacccccatatcaccaagggcaaaaatgg 1618  
Db 9359 CTCACCCCGAGTAGATATTTTGCATAACATTTGACCCCATATACACCAAGGCAAAAAATGG 9300  
QY 1619 ctctctgggcagcagcgtatgggatctggaacactcgaatcccaagtcagccatcagagtga 1678  
Db 9299 CTCCTGGGCAGCAGGCTATGGGATCTGGAACTCTGGAACACTGCAATCCAGTCAGCCATCAGAGTGCA 9240



```
vector_side:left
misc_feature 968..2133 /note="assembly_fragment"
misc_feature 2234..3987 /note="assembly_fragment"
misc_feature 4088..6971 /note="assembly_fragment"
misc_feature 7072..10879 /note="assembly_fragment"
misc_feature 10980..15646 /note="assembly_fragment"
misc_feature 15747..25208 /note="assembly_fragment"
misc_feature 25309..33458 /note="assembly_fragment"
misc_feature 33559..46617 /note="assembly_fragment"
misc_feature 46718..60565 /note="assembly_fragment"
misc_feature 60666..92690 /note="assembly_fragment"
misc_feature 92791..122716 /note="assembly_fragment"
misc_feature 122817..149492 /note="assembly_fragment"
misc_feature 149593..165147 /note="assembly_fragment"
clone_end:T7
vector_side:right
BASE COUNT 52130 a 29805 c 30056 g 51853 t 1303 others
ORIGIN

Query Match 66.9%; Score 1515; DB 2; Length 165147;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1529; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

Qy 719 aaagtatcagatcacacacgacttcaacattctatcataagacactcaccacccaactg 778
Db 101883 AAGGTATCAGATACACACCGNCTTCACATTCATCATATAGACCTCCCAACCCACTG 101824

Qy 779 ttaccttggaatgcacacctcacctcctcagaacctgaagaggttgatattcaacgca 838
Db 101823 TTACCTCTGGCAATGCCACCTACCTACCAAACTGAAGAGGTGGATATTCACGCA 101764

Qy 839 tatgtcgaataatgcacttgggtttttacagaaagaatgcacccaccagaaggg 898
Db 101763 TATGTCGGAAATGGCACTTGGGTTTTTACAGAAAAGAAATGCATGCCACCAAGAGG 101704

Qy 899 atttgataaccttttgggtcccttttgggaagtgggtgattactatcacactacaaatg 958
Db 101703 ATTTGATACCTTTTGGTTCCTTTTGGGAAGTGGGGATTAATACACTACAAATG 101644

Qy 959 tgacgtcctcggagatgtgtggtctgaacttgatgaacacgacaatgtcgtgggacta 1018
Db 101643 TGACAGTCTGGGATGTGTGGCTATGACTTGTATGAAAACGACAATGCTGCTGGGACTA 101584

Qy 1019 tgacaatggcatatactccacagaatgtacactcagagagtacagcaaatcttagcttc 1078
Db 101583 TGACAAATGGCATATACTCCACACATGTACACTCAGAGGTACGAAATCTTTAGCTTC 101524

Qy 1079 ccaataccccacaagccttatatttttatatttccttcatcaagctgttcattccacct 1138
Db 101523 CCATAACCCACAAAGCCTATATTTTATATATATGCGCTATCAAGCTGTTCAATCA 101464

Qy 1139 gcaagctcctgcaggtatttctgaacatacogactccattatcaacataaacagagagag 1198
Db 101463 GCAAGCTCTGGCAGGTATTTTCGAACACTACCGATCCATTATCAACATAAACAGGAGG 101404

Qy 1199 atatgctgccatgcttccctcttagatgaagcaatcaacacgtgacattggctctaaa 1258
Db 101403 ATATGCTGCCATGCTTTCCTGCTTAGATGAAGCAATCAACACGCTGACATGGCTCTAAA 101344
```

```
Qy 1259 gacttatgttttctataacaacagcattatcatcttactcttcagataaatgggtgcagcc 1318
Db 101343 GACTTATGTTTCTATACACACAGCATTTATCATTTACTTCTCAGATATATGGTGGCCAGCC 101284

Qy 1319 tacggcaggaaggagtaactggcctctcagagggtgacaaagaaacatatattgggaaggagg 1378
Db 101283 TACGGCAGGAGGAGTAACGTGCCCTCTCAGAGGTAGCAAGGAACATATTGGGAAGGAGG 101224

Qy 1379 gatccggcgttaggtctttgtgcataagcccactcttgaaaaaacaagggaacagtgtgtaa 1438
Db 101223 GATCGGGCTGTAGGCTTTGTGCATAGCCCCACTTCTGAAAAACAAGGGAACAGTGTGTA 101164

Qy 1439 ggaacctgtcacatcactgactgctgatacccaactctattctcactggctgaagagcacat 1498
Db 101163 GGAACCTTGTGCACATCACTGACTGCTGTACCCCACTCTCATTTTCACCTGGCTGAAGGACAGAT 101104

Qy 1499 tgatgaggacattcaactagatggctatgatatctgggagacacataaagtgaagggtcttcg 1558
Db 101103 TGATGAGGACATTCAACTAGATGGCTATGATATCTGGGAGACCATAAGTGAAGGCTCTTCCG 101044

Qy 1559 ctccaccgagtagatattttgtcataaacatgaccccaataatacaccaagggaacaaatgg 1618
Db 101043 CTCACCCCGAGTAGATATTTTGCATAAACATTGACCCCAATATACACAAAGGCAAAAATGG 100984

Qy 1619 ctctcgggcagcaggtctatgggatctggaacactgcaatccagtcagcagcagagtgca 1678
Db 100983 CTCCTGGCAGCAGGCTATGGGATCTGGAACACTGCANTCCACTCAGCCATCAGAGTGCA 100924

Qy 1679 gcactggaaattgcttacagaaaacctcctggctacagcgcactgggtgggtcccccctcagctttt 1738
Db 100923 GCATCGGAAATTCCTTACAGGAAATCCTTGGCTTACAGCGACTGGTCCGCCCTCAGTCTTT 100864

Qy 1739 cagcaacctgggacgaaaccgggtggcacaatgaacgagatcacctcgtcaactggcaaaag 1798
Db 100863 CAGCAACCTTGGGACCGAACCGGTGGCAATGAACGGATTCACCTTGTCAACTGGCAAAAG 100804

Qy 1799 tgatgggttttcaacatcacagccgaccccatatgagaggtgggacctatctaaacaggtca 1858
Db 100803 TGTATGGCTTTTCAACATCACAGCCGACCCCATATGAGAGGTGGGACCTATCTTAACAGGTA 100744

Qy 1859 tccaggaatcgtgaagaagctcctcagggagctctcacagtttcaaaaaactgcagtgcc 1918
Db 100743 TCCAGGAATCGTGAAGAAAGCTCCTACGGAGGCTCTCACAGTTCACAAACCTGCAGTGCC 100684

Qy 1919 ggtcaggtatcccccaagaagccccagagaagtaaacctcaggtcgaatgaggggtctaggg 1978
Db 100683 GGTCAAGGTATCCCCCAAGACCCCAAGTAACCTTAGCTCAATGAGGGGTCTGGGG 100624

Qy 1979 accatgggtatagagaggaacccaagaaaaaagaagccaagcaaaaatcagggctgagaaaaa 2038
Db 100623 ACCATGGTATAAAGAGGAACCAACGAAAAAGAAAGCAAGCAAAATCAGGCTGAGAAAA 100564

Qy 2039 gcaaaagaaaagc-aaaaaaagaagaacacagacagagaagcagtcctcaggttcaacttg 2097
Db 100563 GCAAAAGAAAAGCAAAAAAAGAAAGAACACAGCAGAAAGCAGTCTCAGGTTTCAACTTG 100504

Qy 2098 ccattcaggttacttgtgataagcacaaatatcttcgtttgtttaaactttaalca 2157
Db 100503 CCATTCAAGTGTACTTGTGGATAAGCAAAATATTTCTGTTTGTAAACTTTAAATCA 100444

Qy 2158 gttctattctttcatctgttctcctaggtaaacacagcaaatgttggctcgataatcgcgtg 2217
Db 100443 GTTCTTATCTTTCATCTCTTTCCTAGGTAAACCAAGCAAAATTTGGCTCGATATAATATCGCTG 100384

Qy 2218 gcctaagcgtcagggctgttttctcatgctgtgccac 2252
Db 100383 GCCTAAGCGTCAGGCTTGTTCATGCTGTGCCAC 100349
```

RESULT 5  
AC091322/c AC091322 232951 bp DNA linear HTG 11-JUN-2001  
LOCUS Mus musculus clone RP23-60D3, WORKING DRAFT SEQUENCE, 5 unordered  
DEFINITION

```

pieces.
AC091322
AC091322.2 GI:14336533
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 232951)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-60D3
Unpublished
2 (bases 1 to 232951)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A.,
Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K.,
Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., McPheeters,R., Meldrum,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S.,
Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (15-APR-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 11, 2001 this sequence version replaced gi:13625486.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L13252
Center clone name: 60_D_3
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 231176 bases at least Q40
Consensus quality: 231965 bases at least Q30
Consensus quality: 232319 bases at least Q20
Insert size: 210000; agarose-fp
Insert size: 232551; sum-of-contigs
Quality coverage: 10.7 in Q20 bases; agarose-fp
Quality coverage: 9.6 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 53: contig of 53 bp in length
* 54 153: gap of 100 bp
* 154 2590: contig of 2437 bp in length
* 2591 2690: gap of 100 bp

```

```

* 2691 11774: contig of 9084 bp in length
* 11775 11874: gap of 100 bp
* 11875 100524: contig of 88650 bp in length
* 100525 100624: gap of 100 bp
* 100625 232951: contig of 132327 bp in length.
FEATURES
    source
        1..232951
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone="RP23-60D3"
            /clone_lib="RPC1-23 Female Mouse BAC"
            1..53
                /note="assembly_fragment"
                clone_end:SP6
                vector_side:left"
                154..2590
                    /note="assembly_fragment"
                    2691..11774
                        /note="assembly_fragment"
                        11875..100524
                            /note="assembly_fragment"
                            100625..232951
                                /note="assembly_fragment"
                                clone_end:T7
                                vector_side:right"
BASE COUNT 70795 a 45474 c 46842 g 69434 t 406 others
ORIGIN

Query Match      50.1%; Score 1134.2; DB 2; Length 232951;
Best Local Similarity 84.7%; Pred. No. 0;
Matches 1303; Conservative 0; Mismatches 218; Indels 18; Gaps 2;

Qy 712 ttactgaaagtatcagatacacacccggacttaacattctatcataagacctaccaac 771
Db 147754 TTCCTCAAGAGGTATCAGATACACACGGGACTTCAGCATTTCTATCATTAGACCTACCCCAAC 147695

Qy 772 ccaactgttacctctgtgacaatgccacctaccctcagaaactgaagaggttgatatt 831
Db 147694 CCAACTGCTACTCTCGACAAATGCACCCCTACCTCAGAACGTAAGAGAGTGGCTATT 147635

Qy 832 caacgcatatggtcgaaaaatggcacttgggttttttacagaaaaaatgcataccacca 891
Db 147634 CAACCCATATGTTGGGAAGTGGCACTTGGGATTTTACAGAAAAAGACTGCATGCCACCA 147575

Qy 892 gaagagatttgataccttttttggttcccttttgggaagtggggttactatatacaact 951
Db 147574 AGAGAGGATTTGATACCTTTTTTGGCTCTCTTTTGGGAAGTGGTATTATTATACACACT 147515

Qy 952 acaaatgtgacagtcctctgggatgtggtctatgactttatgaaaacgacaatgctgct 1011
Db 147514 ACAANTGTGACAGTCTCTGGGGTGTGGCTATGACTTGTAGAAAACGACAAATGCAGCTT 147455

Qy 1012 gggactatgacaatggcataatactctccacacagatgtacactcagagatgacagaaatct 1071
Db 147454 GGGACTATGACAATGGGCATCTATTCAACACAGATGTACACGACAGAGTGCAGCAAAATCT 147395

Qy 1072 tagcttccccatacccccaagaagcctatatattttatatattgcctataaagctgttcatt 1131
Db 147394 TAGCCACCATGATCCCCACAAACCTCTATTATATGTGGCTTACCAGCTGTGCCACT 147335

Qy 1132 caccattcaagctcctctggcaggtatttcgaacactaccgatccattatcaacataaaca 1191
Db 147334 CACCACCTTCAGAGCCCTCGCAGGTACTTTGAACACTACAGATCCATTATCAACATAAACA 147275

Qy 1192 ggaggagatagtgcacatgctttctctgcttagatgaagcaatcaacaacgtgcattgg 1251
Db 147274 GGAGGAGTACCGAGCCCATGCTCTCTCTGCTGGGAAGCAATCCACAATGTGACCCCTGG 147215

Qy 1252 ctctaaagactattggtttcttataaacaacagcattatcttactcttcaagaataatggtg 1311
Db 147214 CCCTAAAGAGGTATGGTTTCTATATAACAATAGCATTTATCTTACTCTCTCAGATAATGGTG 147155

```





Db	319	CTCAGCTTTTCAGCAACCTGGACCGAATCGGTGGCGAATGAACGGATCACCTTGTCAA	378
Qy	1789	ctggcaaaagtgtatggttttcaacatcacagccgacccatataagaggtgacccat	1848
Db	379	CTGGCAAAAGTGTATGGCTTTTCACATCACAGCCGACCCATATAGAGGTGGACCTAT	438
Qy	1849	ctaacaggtatccagggaatcgtgaagaagctctctacagaggtctcacagttcaacaaaa	1908
Db	439	CTAACAGGTATCCAGGAATCGTGAAGAAGCTCTACGGAGGCTCTCACAGTTCAACAAAA	498
Qy	1909	ctgaagtcgccggtcaaggtatcccccaagaagaccaggaagtaacctaggtcctaatgag	1968
Db	499	CTGCAGTCCCGGTAGGTATCCGCCCAAGACCCGAGAAGTAACCTTAGGTCAATGGAG	558
Qy	1969	gggtctaggaccatggtatagagaggaaacaaagaaaaagaaagcgaagcaaaatcagg	2028
Db	559	GGGTCTGGGACCCTGTTAATAGAGGAACCAAGAAAAAGCAAGCAACCAAAATCAGG	618
Qy	2029	ctgagaaaaaagaaaaagc-aaaaaagaagaagaaacagcagagaaacagtcctcag	2087
Db	619	CTGAGAAAAAGCAAAAGAAAGCAAAAGAAAAAGAAAGAAAGCAAGCAAGCTCTCAG	678
Qy	2088	gttcaacttgccattcaaggttacttcttgataagcacaatatctctgttggttcaa	2147
Db	679	GTTCAACTTGCCATTCAGGTGTTACTTGTGATTAAGCAAAATATTTCTGTTGGTTAA	738
Qy	2148	actttaatcagttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct	2207
Db	739	ACTTTAATCAGTCTTATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	798
Qy	2208	aatatcgtgctgaagcgtcaggtgttcttcttcttcttcttcttcttcttcttcttct	2252
Db	799	AATATCGTGGCTAAGCTCAGGCTTCTTTTCATGCTGTGCCAC	843
RESULT	7		
AC013692/c			
LOCUS		157043 bp	DNA linear HTG 07-DEC-2000
DEFINITION		Homo sapiens clone RP11-211L8, *** SEQUENCING IN PROGRESS ***	45
ACCESSION	AC013692		
VERSION	AC013692.5	GI:11597077	
KEYWORDS	HTG; HTGS_PHASE1.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
JOURNAL	1 (bases 1 to 157043)		
REFERENCE	2 (bases 1 to 157043)		
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,		
	Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhalter, B.,		
	Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A.,		
	Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,		
	Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,		
	Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,		
	Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,		
	Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,		
	McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,		
	Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,		
	Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,		
	Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,		
	Tesfaye, S., Tirrell, A., Vassiliou, H., Vo, A., Wheeler, J., Wu, X.,		
	Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.		
TITLE	Direct Submission		
JOURNAL	Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome		
	Research, 320 Charles Street, Cambridge, MA 02141, USA		
COMMENT	On Dec 7, 2000 this sequence version replaced gi:10280848.		
	All repeats were identified using RepeatMasker:		
	Smit, A.F.A. & Green, P. (1996-1997)		

http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L4053  
Center clone name: 21\_L\_8  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 45 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1 924: contig of 924 bp in length  
\* 925 1024: gap of 100 bp  
\* 1025 1800: contig of 776 bp in length  
\* 1801 1900: gap of 100 bp  
\* 1901 2800: contig of 900 bp in length  
\* 2801 2900: gap of 100 bp  
\* 2901 3628: contig of 728 bp in length  
\* 3629 3728: gap of 100 bp  
\* 3729 4582: contig of 854 bp in length  
\* 4583 4682: gap of 100 bp  
\* 4683 5763: contig of 1081 bp in length  
\* 5764 5863: gap of 100 bp  
\* 5864 6992: contig of 1129 bp in length  
\* 6993 7092: gap of 100 bp  
\* 7093 8691: contig of 1599 bp in length  
\* 8692 8791: gap of 100 bp  
\* 8792 10237: contig of 1446 bp in length  
\* 10238 10337: gap of 100 bp  
\* 10338 12115: contig of 1778 bp in length  
\* 12116 12215: gap of 100 bp  
\* 12216 14480: contig of 2265 bp in length  
\* 14481 14580: gap of 100 bp  
\* 14581 16200: contig of 1620 bp in length  
\* 16201 16300: gap of 100 bp  
\* 16301 17801: contig of 1501 bp in length  
\* 17802 17901: gap of 100 bp  
\* 17902 19560: contig of 1659 bp in length  
\* 19561 19660: gap of 100 bp  
\* 19661 21061: contig of 1401 bp in length  
\* 21062 21161: gap of 100 bp  
\* 21162 22798: contig of 1637 bp in length  
\* 22799 22898: gap of 100 bp  
\* 22899 25224: contig of 2326 bp in length  
\* 25225 25324: gap of 100 bp  
\* 25325 27369: contig of 2045 bp in length  
\* 27370 27469: gap of 100 bp  
\* 27470 30262: contig of 2793 bp in length  
\* 30263 30362: gap of 100 bp  
\* 30363 32450: contig of 2088 bp in length  
\* 32451 32550: gap of 100 bp  
\* 32551 34415: contig of 1865 bp in length  
\* 34416 34515: gap of 100 bp  
\* 34516 37321: contig of 2806 bp in length  
\* 37322 37421: gap of 100 bp  
\* 37422 40476: contig of 3055 bp in length  
\* 40477 40576: gap of 100 bp  
\* 40577 50235: contig of 9659 bp in length  
\* 50236 50335: gap of 100 bp  
\* 50336 53457: contig of 3122 bp in length  
\* 53458 53557: gap of 100 bp  
\* 53558 56770: contig of 3213 bp in length  
\* 56771 56870: gap of 100 bp  
\* 56871 59626: contig of 2756 bp in length  
\* 59627 59726: gap of 100 bp  
\* 59727 62247: contig of 2521 bp in length















```
Qy 1500 gatgaggacattcaactagatgctatgatatactctggagaccataaagttagggtcttcgc 1559
Dy 102455 TCACGAGCGGATGGCTAGATGGCTACGACGCTGTGCGCGCCATCAGCGAGGCGCGGCC 102514
Qy 1560 tcaccccgagtagatttggataataacttgaccatatacaccaaggcaaaaatggc 1619
Dy 102515 TCACGAGCGGAGATGCTGCACCAACATTTGACCCACTCTACAAGCATGCCAGCATGCG 102574
Qy 1620 tcttggcgagcgatggtggtctggaacactgcaatccactgacccatgagatgcag 1679
Dy 102575 TCCTGGAGGCGGCTTGGCATCTGGAAACCGCGGTGCGACGCTGCGCATCCGCGTGGGT 102634
Qy 1680 cactggaattcttacaggaaatcctggctacagcagctgggtcccccctcagctcttc 1739
Dy 102635 GAGTGGAAAGCTGCTGACGAGGAGACCCCGCTATGCGGATGCGATCCACCGCAGACACTG 102694
Qy 1740 aqcaacctgggacgaacacgggtggcaccaatgaacgggataccctctcaactggcaaaagt 1799
Dy 102695 GCCACCTTCCCGGGTAGTGGTGGGAACCTGGAAACCAATGGCCAGTG- --TCCGCCAGGCC 102751
Qy 1800 gtatggcttttcaacatcacacgacccatgacccatgagaggtggacctatctaacaggat 1859
Dy 102752 GTGTGGCTTCAACATCAGTCTGACCCCTATGAAACGGGAGGAGCTGGCTGGCCAGCGG 102811
Qy 1860 ccaggaatgtgaagaactcctcagcgggtctcacagttcacaacaaactgcagtgcg 1919
Dy 102812 CCTGATGTGGTCCGACCCCTGTGCTGCGCTGCGCGGAATATAACCGCACAGCCATCCG 102871
Qy 1920 gtcaggtatcccccaagacaccccaagaagtaacctagctcaatggggggtctaggg 1979
Dy 102872 GTACGCTACCCAGCTGAGAAACCCCGGCTCATCTCTGACTTTAATGGGGGTCTTGGGG 102931
Qy 1980 ccattgtagagagagaaaccaaagaaaaagaaagca 2019
Dy 102932 CCTCGGCCAGTGATGAGGAAGAGAGGAGGAGGAGGA 102971

RESULT 13
AF109924 2805 bp mRNA linear INV 05-MAY-2000
LOCUS Helix pomatia sulfatase 1 precursor (SULF1) mRNA, complete cds.
DEFINITION Helix pomatia sulfatase 1 precursor (SULF1) mRNA, complete cds.
ACCESSION AF109924
VERSION AF109924.1 GI:6863175
KEYWORDS .
SOURCE Roman snail.
ORGANISM Helix pomatia
Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata;
Stylommatophora; Helicacea; Helicidae; Helix.
REFERENCE 1 (bases 1 to 2805)
AUTHORS Wittstock,U., Fischer,M., Svendsen,I. and Halkier,B.A.
TITLE Cloning and characterization of two cDNAs encoding sulfatases in
the Roman snail, Helix pomatia
JOURNAL IUBMB Life 49, 71-76 (2000)
MEDLINE 20233298
REFERENCE 2 (bases 1 to 2805)
AUTHORS Wittstock,U., Fischer,M., Svendsen,I. and Halkier,B.A.
TITLE Direct Submision
JOURNAL Submitted (27-NOV-1998) Plant Biology, The Royal Veterinary and
Agricultural University, Thorvaldsensvej 40, Frederiksberg C DK
1871, Denmark
FEATURES
source
1..2805
/organism="Helix pomatia"
/db_xref="taxon:6536"
1..2805
/gene="SULF1"
149..1660
/gene="SULF1"
/codon_start=1
/product="sulfatase 1 precursor"
/protein_id="AAF30402.1"
/db_xref="GI:6863176"
/translation="MCKCLLVLIATITACAVADOSSASAGTRQDAGOPNIVFVLADDF
```

```
GFHDVGYHGSIIHTPTLDALSASGVRLNYYVQPICTPTRSOLMSGRYOIHTGLQHLGI
INSCOPNALPNDSPITLADKLKESGYATHMVGKHLGYKQEVLPNMRGDTYVGYLNA
AEDYENHNVPMPQVRYLDLRRNGNVPVRNETGOYSAHLFTGKDAIDVQSHNTPSKPLFLY
LAYOSVHAPLEVPYEHKYNITDKNRRTFAGWVSALDEGVANITQALKDKGLWNNT
VLIFSTDNGGQIHAGGNNYPLRGKASLWEGFHVSGGALKRSAGVSKGLIHVS
DMFTPLVTLAGNLNGTKPLDFNQMDTISNETPSPREILLHNIDILYLPQKGVPLYSN
TWDRVRAAIRVGDYKLTJGPDGNGSWPPDGHLYFVPEIQESAANKNVLNFINITADP
NEHNDLSSEKPLEVLRLLQILVQENNTAVPPRYPADPDCRDPALHGDVWGPEW"
BASE COUNT 871 a 611 c 656 g 667 t
ORIGIN
Query Match 17.6%; Score 398.2; DB 3; Length 2805;
Best Local Similarity 56.9%; Pred. No. 3.1e-108;
Matches 824; Conservative 0; Mismatches 593; Indels 30; Gaps 4;
Qy 541 cctccacgcccccatctcatcttcactagcggatgatcaggagatttagagatgtgggtt 600
Dy 240 CCGGTCAACCCCAACATTTGTTTGTGTAGCAGACGACTTCGGTTTCCATGACGTCGGTT 299
Qy 601 accacggatctgagattaaacacactactcttgacaagctcgtccggaaggattaaac 660
Dy 300 ACCATGGTTCGAGATCCACACCCCAACACTTGTATGCGCTGTCGCCACCGCGTCAGGC 359
Qy 661 tggaaactactatgtccagcctatttgcacaccatccaggagtcagtttattactggaa 720
Dy 360 TGGAGAACTACTACGTGCGAGCCGATTTGCACCCCGCAGGAGTCAGCTCATGTCCGGGA 419
Qy 721 agtatcagatacacacccggacttcaacattctatcataagacctaccacccaactgtt 780
Dy 420 GATATCAGATACATATCGGCTTCAACATGGTATTATTAAATAGTTGTCAGCCAAAGCGTC 479
Qy 781 tacctctggacaatgccacctactcagaaactcagaagaggttgattatcaacgacata 840
Dy 480 TTCCAAATGACAGTCCGACTTTGGCAGATAAACTGAAAGAACTGGCTGCGCCACACACA 539
Qy 841 tggtcgaaaaatggcacttgggtttttacagaaaagatgcatgccccacagaagagat 900
Dy 540 TGGTTGGCAAGTGGCACCTAGGATTCTACAAACAGAACTACTTGCCTTGGAAACCGAGGT 599
Qy 901 ttgataccttttttgggtcccttttgggaagtggggattactatcacacatacaaatgtg 960
Dy 600 TCGACACTTACTTTGGTTATCTTAATGTCTGAGGACTACTTTAACCAACAATGTACCTT 659
Qy 961 acagtccttgggagtgtgtgctatgactgtatgaaaacgacaactgctgctgggactatg 1020
Dy 660 GGAGACAAAGTG-----CGCTACCTTGACCTCAGAGATACAAACGGACCCGTGAGAAAG 713
Qy 1021 acaatggcatatactccacacagatgtacactcagagagacagcaaatcttagcttccc 1080
Dy 714 AAACCTGGGCAGTATTTCAGCTCACCTGTTCACCGGCAAGCCCATTTGATGTTGTACAGTCA 773
Qy 1081 ataaccacacaaagcctatattttatataattgctctatacaagctgttctatcaccactgc 1140
Dy 774 ACAACACGCTCCAAGCCGCTATTCTTGTACCTGGCTACCAGTCCGTCGACGCCGCCACTTG 833
Qy 1141 aagtccttggcaggtatttccgacactaccgattccattatcaacataaacaagagagat 1200
Dy 834 AGGTCCAGAAAAGTACGACACACAGATATAGAAACATCACACACAAAACAGCGCCACTT 893
Qy 1201 atgtgcccattgcttctgcttagatgaagcaatcaacaacgtgacattggctctaaaga 1260
Dy 894 TCGCGGGTATGCTCTCAGCCCTGGATGAGGGGTAGCCCAACTTAACCCAGGCACCTCAAAG 953
Qy 1261 cttagtggtttctatacaacaacagcattataattactcttcagataatggtggccagccta 1320
Dy 954 ACAAGGGACTCTGGAAACAATACTGCTGTGATATTTTCTACAGACAATGGCGGGCAGATAC 1013
Qy 1321 cggcagggagtagtaactggcctcagagtagcaaaagaaacatatgggaagagagga 1380
Dy 1014 ATGCTGGAGGGACAATAATATCCTCTCGTGGCTGGAAGGCATCGGCTGGGAGGTGGAT 1073
Qy 1381 tccgggctgtaggcttgtgtcatgccaccttctgaaacaaaggaagacagtggttaagg 1440
```

```

Db 1074 TTATGGGTGGGTCTGCTGGAGAGCACTGAACCGAGTGGAGCTGTCTAGTAAAG 1133
Qy 1441 aacctgtgcacactcaactgactggtaccacactctcattctcactgctgaagcagattg 1500
Db 1134 GATTGATTCAGCTTCTGATGGTTCCTACTTTAGTAACACTAGCTGGTGGAAATTGA 1193
Qy 1501 atgaggacattcaactagatgcttatgatatctctgggagaccataagtgagggtcttcgct 1560
Db 1194 ATGGAACCAAGCCATTGGATGGTTTAAACCAATGGGATACATAAGCAACGAGACTCCTT 1253
Qy 1561 caccctgagtagatatttgcataaactgacattgaccccatatacaccaaggcaaaaatggct 1620
Db 1254 CGCCGAGGGAATCCTCTCCACAACTGACATCTCTGTACCCCA--AAAAGGTGTGCC 1311
Qy 1621 cctgggagcagcaggctatgggtctggaacactgcaactcagtcagcaccatcagagtgccagc 1680
Db 1312 CCTGTACAGCA-----ACACTGGGACACGAGGGTCAAGGCAGCTATCAGAGTTGGGG 1364
Qy 1681 actggaattgcttacaggaatactgctgctacagcagctggtggtcccccctcagttcttca 1740
Db 1365 ACTACAAGCTGATTACCGGGACCCAGGCAATGGCAGCTGGGTGCTCCACCTGATGGC- 1423
Qy 1741 gcaacctgggaccgaacccggtggcacaatgaacgggatcacctcgtcaactggcgaagaagt 1800
Db 1424 -----CACCTGTATTTTGTACCTGAAATCCAAAGATTCGGCTCGCAAAAACG 1469
Qy 1801 tatgcttttcaacatacagccacccatgatgagagggtgagacctatctaacagggtatc 1860
Db 1470 TGTGCTGTTTAACTACACCGCGACCCAAAGAGCATAAATGATCTGTCAGTGAGAAAC 1529
Qy 1861 caggaatcgtgaagaagctctacgagagctctcacagttcaacaaaactgcagtgccgg 1920
Db 1530 CACTTGAGGTCCTGAGACTGCTGCAAAATACCTGGTCCAGCTTTAAACAATACAGCAGTGCCAC 1589
Qy 1921 tcaggatcccccccaagaagccccagaagtaacccttagctcaatggaggggtctaggagac 1980
Db 1590 CTAGATACCCCGCACACAGACCCAGTGGCGACCCAGCTCTGCATGGTGATGCTCTGGGGAC 1649
Qy 1981 catggtga 1987
Db 1650 CGTGGGA 1656

RESULT 14
S48472
LOCUS      S48472      1939 bp      mRNA      linear      MAM 08-MAY-1993
DEFINITION arylsulfatase B [cats, liver, mRNA, 1939 nt].
ACCESSION  S48472
VERSION    S48472.1  GI:258855
KEYWORDS   .
SOURCE     cat liver.
ORGANISM   Felis catus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
            1 (bases 1 to 1939)
            Jackson,C.E., Yuhki,N., Desnick,R.J., Haskins,M.E., O'Brien,S.J.
            and Schuchman,E.H.
            Feline arylsulfatase B (ARSB): isolation and expression of the
            cDNA, comparison with human ARSB, and gene localization to feline
            chromosome A1
            Genomics 14 (2), 403-411 (1992)
            93052342
JOURNAL    GenBank staff at the National Library of Medicine created this
MEDLINE    entry [NCBI gibbsq 117975] from the original journal article.
REMARK     This sequence comes from Fig. 2.
            Map location: A1.
            Location/Qualifiers
            1..1939
            /organism="Felis catus"
            /db_xref="taxon:9685"
            1..1939
            /gene="aryl sulfatase B, ARSB"
```

```

CDS
4..1611
/gene="aryl sulfatase B, ARSB"
/note="This sequence comes from Fig. 2; ARSB"
/codon_start=1
/product="aryl sulfatase B"
/protein_id="AAB23941.1"
/db_xref="GI:258856"
/translation="MGRRAASLPGRSPRRLPLPGVLPDLLLRLLSRPGAGAGAD
RPHLFVFLADLIGWINDVSPHGSINRTPHLDELAAGVLLDNYTQPLCTPSPRSQLLT
GRYQIHGTGLQHOIIPQPCVPLDEKLLPQLKEAGYTHMVGKHLGMVSRKCLPT
RRGFDTVGYLLGSEEDYSHERCALIDSLNVTBCALDFRDEQVATCYKMYSPNIET
ERATALTSHPPKPLFLYLALQSVHEPQVPEYLKPYDFIQDKNHYHAGWVSLMD
EAVGNITAAKSLHGLMNTVIFSTDPNGQVPLAGNNWPLRGKWSLWEGIRGVGF
ASPLLKOKGVKNRELIHISDLPTLVKLARGSTKPLKDFVWKTSISGSPSPRKE
LLHNIDPNFVIDISPCGKSLAPAKDDSSHPAYLAFNTSLHAAIRHGNNKLLTGYPGCG
CWPPPPSPYNDSAITPSSDPPTKTLWPFIDIDQDPERHDLSDRYPHIVEQLLRSLRQFYH
KHSVPVHFPAQDPRCDPKGTGANGPWV"
BASE COUNT      450 a 552 c 503 g 434 t
ORIGIN
Query Match      16.7%; Score 379; DB 4; Length 1939;
Best Local Similarity 56.4%; Pred. No. 1.7e-102;
Matches 830; Conservative 0; Mismatches 575; Indels 66; Gaps 4:
Qy 546 cagcccatctcatttctcctcctagcggatgatcagggatttagagatgtgggtaccac 605
Db 139 CGCGGCATCTCGTCTTCTGTTGGCGGACGACCTGGCTGGAAGCAGCTGAGCTTCCAC 198
Qy 606 ggatctgagataaaacacactactcttgacaagctcgtgcgaaggaggttaaacctgag 665
Db 199 GGCTCGAATATCCGACACGCCGACCTGGACGAGCTGGCAGCGCGGGGTGCTCTGGAC 258
Qy 666 aactactatgtccagcctatttgcacaccatccaggagtcagttatttaccaggaaagt 725
Db 259 AACTACTACAGCAACCGCTGTGCACGCCATCGCGGAGCCAGCTGCTCACC GGCGCTAC 318
Qy 726 cagatcacacacggacttcaacattctatcaagaacctaccaccaacactgtttacct 785
Db 319 CAGATCCACACAGGTTTACACACCAATAATCTGCGCCCTGTTCAGCCCGAGCTGCTCCT 378
Qy 786 ctggacaatgcacccctaccctcagaacctgaaggaggttgatattcaacgcataatgtc 845
Db 379 CTGGATGAAAACCTCTGCCCCAGCTTCTAAAGAAGCAGGCTACACTACCATATGTC 438
Qy 846 ggaataatggcactgggtttttacagaaaagaatgcagccaccacagagagatttgaat 905
Db 439 GGAATAATGGCACCTGGGAATGTACCGGAAAGAATGCTTCCAACCCCGGAGGATTTGAT 498
Qy 906 acctttttgttcccttttgggaagtgggattactatacacactacacaaatgtgacagt 965
Db 499 ACTTACTTTGGATATCTCTTAGGTAGTCAAGATTACTATCCCATGAGCGCTGTGCATTA 558
Qy 966 cctggga-----tgtgtggtctatgactgtatgtaaaacgacaatcgtgcc 1010
Db 559 ATTGACAGCTCGAATGTGCACACGATGTGCTCTGTGATTTTCGAGATGGGGAACAGGTGCA 618
Qy 1011 tggactatgacaatggcatactaccacacagatgtacactcacagagtagacagcaaatc 1070
Db 619 ACAGATATATAAAAT---ATGTATTTCGACAAACATATTTACTGAAAGAGCTACACCCCTC 675
Qy 1071 ttacttccccataacccccacaaagcctatatatttttatttgcctatcaagctgttcat 1130
Db 676 ATAAC TAGCCATCCACCCCGAGAGCGCTCTGTTTCTCTACCTTGTCTTTCAGTCTGTCCAC 735
Qy 1131 tcaccactgcaagctcctggaagtgatttcgaacactaccatccatccatatacaacaataac 1190
Db 736 GAACCCCTTCAGGTCCCTCGAGAGTACCTGAAACCCCTACGACTTTTATCCAGATAAGAAAT 795
Qy 1191 aggaggagatgatgtcccatgctcttccctgcttagtagaagcaatcaacaacgtgacattg 1250
Db 796 AGGCATTACTATGCAGGAATGCTGTCTTATGATGAAGAAGCAGTGGGAATGTCACAGCA 855
```



```
Db 1161 AAGTTGCAACAGGATATAAAAT---ATGTATTCAACAACAAATATTCAACAAAAGGGCTA 1217
Qy 1063 agcaaatcttagctccataaccccccaaaagccctatatattttatatattgcctatcaag 1122
Db 1218 TAGCCCTCATAACTAAACCATCCACGAGAAGCCTCTGTTCTTACCTTGCTCTCCAGT 1277
Qy 1123 ctgttcattccactccaagctcctcctggcaggtattttogaacactaocgattccattatca 1182
Db 1278 CTGTGCATGAGCCCTTTCAGGTCCTCGAGTAATACTTGAAGCCATATGACTTTATCCAAG 1337
Qy 1183 acataaacagagagagatatgtcccatcttccctctcttagatgaagcaatcaacaacg 1242
Db 1338 ACAAGACAGGCATCACTATGACGAATGGTCCCTTATGGATGAAGCAGTAGGAAATG 1397
Qy 1243 tgacattggctctaaagacttatgtttctataaacaacagcatattatcttactcttcag 1302
Db 1398 TCACTGCGAGCTTTAAAGCAGTGGGCTCTGGAACAACACACGCGTGTTCATCTTTTCTACAG 1457
Qy 1303 ataagtgtgcccactacggcagggaggttaactggcctctcagaggttagcaaaagaa 1362
Db 1458 ATAACGGAGGCGACACTTTGGCAGGGGTAAATAACTGGCCCTTCGAGGAAGAAAATGGA 1517
Qy 1363 catattgggaagggggtccggctctaggtcttctgtgcatagccactctctgaaaaaca 1422
Db 1518 GCCTGTGGGAAGGAGCGTCCGAGGGTGGGCTTTGGCAAGCCCTTGCTGAAGCAGA 1577
Qy 1423 agggaaacagtgttaaggaaacctgtcacatcactgactggttacccactctctattcac 1482
Db 1578 AGGGCGTGAACAACGGGAGCTCATCCACATCTCTGACTGGCTGCCACACACTCGTGAAGC 1637
Qy 1483 tggctgaaggacagattgatgagacattcaactagatggctatgatctctgggagacca 1542
Db 1638 TGGCCAGGGGACACACCAATGGCAAAAGCCTCTGGATGGCTTGACATGTGGAACCA 1697
Qy 1543 taagtgaaggctctcgctcccccagtagatatatttgataaactgaaccccatatac- 1601
Db 1698 TCAGTGAAGGAAGCCCATCCCCAGAAATTGAGCTGCTGCATAAATATTGACCCGAACTTCG 1757
Qy 1602 -----accaaggcaaaaaatggctcctctgggag 1629
Db 1758 TGGACTCTTACCGTGTCCAGGAACAGCATGGCTCCAGCAAGGATGACTTCTCTTTC 1817
Qy 1630 caggctatgggactctggaacactcaatcccagtcagccatcagagtgagcactggaaat 1689
Db 1818 CAGAAATATTACGCTTTAAACACATCTGTCCATGCTGCAATTAGACATGGAATTTGGAAC 1877
Qy 1690 tgettacaggaatacctggctacagcactgggtcccccctcagctctttcagcaacctgg 1749
Db 1878 TCCTCAGGGGTACCCAGGCTGTGTTACTGGTTCCCTCCACCGCTCAATACAAATGTTT 1937
Qy 1750 gaccgaaccggtgggcacaatgaacggatcacctcgtcaactggcaaaagtgtatggcttt 1809
Db 1938 CTGAGA-----TACCTTCATCAGACCCCAACCAAGACCCCTCTGGCTCT 1982
Qy 1810 tcaacatacagcgcgaccctatatagaggggtggacctatcttaacagggtatccaggaatcg 1869
Db 1983 TTGATATTGATCGGGACCCCTGAAGAAAGACATGACCTGTCCAGAGAAATATCTTCACATCG 2042
Qy 1870 tgaagaagctcctcagggaggtctcacaagtccaacaaactgcagtgccggtcaggtatc 1929
Db 2043 TCACAAAGCTCTGTCCCGCTACAGTTCTACCAATAAACACTCAGTCCCGCTGACTTCC 2102
Qy 1930 ccccaagagccccagagtagaacctagggtcaatggaggg 1970
Db 2103 CTGCACAGGACCCCGCTGTGATCCCAAGGCCACTGGGGTG 2143
```

Search completed: July 23, 2002, 17:14:59  
Job time: 7805 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 14:59:59 ; Search time 2394.35 Seconds  
(without alignments)  
12773.446 Million cell updates/sec

Title: US-09-495-823-8  
Perfect score: 2266  
Sequence: 1 caccgctcgccacgcgc.....tgccacctgggtccgaattc 2266

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estnu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_gss:\*
- 13: em\_gss\_hum:\*
- 14: em\_gss\_inv:\*
- 15: em\_gss\_pln:\*
- 16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	703.8	31.1	817	10	BG680752
2	646.2	28.5	683	10	BF307503
3	603.8	26.6	615	9	AW474222
4	512.4	22.6	625	12	AZ815176
5	511.2	22.6	648	9	BB520951
6	506.6	22.4	700	9	BB319509
7	393.8	17.4	463	9	AA159953
8	368.8	16.3	414	10	W81455
9	361.4	15.9	459	10	BF653682
10	341.8	15.1	547	10	BG661302
11	333.2	14.7	344	10	W81485
12	331.2	14.6	458	10	R43179
13	307.4	13.6	325	10	U46280
14	284.6	12.6	1157	10	BM461300
15	262	11.6	628	10	BE553565
16	261.4	11.5	625	10	BI104707
17	245.8	10.8	1032	10	BF979627

c	18	241.4	10.7	735	10	BG772761
	19	226.6	10.0	608	9	AA123795
	20	215.2	9.5	590	9	AI386106
	21	209	9.2	566	9	AA124434
	22	199.4	8.8	370	10	BG346926
	23	198.6	8.8	274	10	BG194726
	24	198	8.7	407	9	AA159952
c	25	196.2	8.7	579	9	AU055665
c	26	196.2	8.7	579	9	AU055666
c	27	192	8.5	775	9	AI327376
	28	191.6	8.5	549	10	BI441091
	29	179.6	7.9	533	9	AV382922
	30	178.8	7.9	942	10	BF160020
	31	176.8	7.8	408	9	AI448125
	32	176.6	7.8	635	9	BB654214
c	33	166.2	7.3	403	9	AA117137
	34	165.8	7.3	423	9	AA358659
c	35	163	7.2	509	9	BB633888
	36	158.6	7.0	542	9	BE014208
	37	157	6.9	849	10	BF696654
	38	148.8	6.6	658	9	BB087596
	39	148.8	6.6	918	10	BF163345
	40	139.4	6.2	987	9	AL666493
	41	122.8	5.4	599	9	AI815689
	42	122.8	5.4	635	10	BE305390
	43	119	5.3	552	10	BF218442
	44	118.8	5.2	412	9	AI791623
	45	118.6	5.2	807	10	BE786803

ALIGNMENTS

RESULT 1

BG680752  
LOCUS 602628445F1 NCI\_CGAP\_Skn4 Homo sapiens cDNA clone IMAGE:4753259 5',  
DEFINITION mRNA sequence.  
ACCESSION BG680752  
VERSION BG680752.1 GI:13912149  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 817)  
AUTHORS NTH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: James Cleaver, M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM0612 row: j column: 12  
High quality sequence stop: 789.  
Location/Qualifiers  
1. 817  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4753259"  
/clone\_lib="NCI\_CGAP\_Skn4"  
/tissue\_type="squamous cell carcinoma"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: skin; Vector: pCMV-SPORT6; Site:1: NotI;  
Site:2: SalI; Cloned unidirectionally. Primer: oligo dt.  
Average insert size 1.5Kb. Library constructed by Life  
Technologies. Note: this is a NCI\_CGAP Library."

BASE COUNT 259 a 193 c 207 g 158 t





```

Qy 1071 ttatgtcccaataccccacaaagcctatatattttatatattgctcatcaagctgttcat 1130
Db 540 TTAGCTTCCCATACCCACAAAGCCTATATTTTATATATATGCTATCAAGCTGTTCAT 599
Qy 1131 tcaccactgaagctcctctggcaggtatttccaaacactaccatccattatcaacaataaac 1190
Db 600 TCACCACATGCAAGCTCCTGGCAGGTATTTCCGAACACTACCCGATCCATTATCAACATAAC 659
Qy 1191 agggaggagatgctgc 1207
Db 660 AGGAGGAGATGCTGC 676

RESULT 3
AW474222/c
LOCUS
DEFINITION xs20h05.x1 NCI_CGAP_Ut-2 Homo sapiens cDNA line EST 24-FEB-2000
similar to SW:ARSB_HUMAN P15848 ARYL SULFATASE B PRECURSOR ;, mRNA
sequence.
ACCESSION AW474222
VERSION AW474222.1 GI:7044328
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 615)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 435.
FEATURES
source
1..615
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2770233"
/clone_lib="NCI-CGAP_Ut2"
/tissue_type="moderately-differentiated endometrial
adenocarcinoma, 3 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.85 kb. Life Technologies catalog #:
11539-012"
BASE COUNT 116 a 151 c 153 g 195 t
ORIGIN

Query Match 26.6%; Score 603.8; DB 9; Length 615;
Best Local Similarity 98.9%; Pred. No. 5.3e-106;
Matches 608; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1437 aaggaacctgtgcacatacactgaactggttacccccactctcttcaactggtgaaggacag 1496
Db 615 AAGAACTGGTGACATCACCACTAGGTGACCCCACTCTCATTTCACTGGCTGAAGGACAG 556
Qy 1497 attgatgaggacattcaactagatgctgatctatctgggagaccataagtgaaggctctt 1556
Db 555 ATTGATGAGGACATTCACACTAGATGGCTATGATATCTGGGAGACCATTAAGTGAGGGTCTT 496
Qy 1557 cgctcaccccgagtagatattttgtcataaacattgacccccatatacaccaaggcaaaaaat 1616

```

```

Db 495 CGCTCACCCCGAGTAGATATTTTGCTAATACATTGACCCCATATACACCAAGCAAAAAAT 436
Qy 1617 ggctcctgggcagcaggtctatgggagctctggaaacactgcaatcccaagcagcatcaagtg 1676
Db 435 GGCTCTCTGGGCAGCAGGCTATGGGATCTGGAACTGCAACTGCAATCCAGTCAGCCATCAGAGTG 376
Qy 1677 cagcactggaattgcttacaggaaatcctggctacagcagctgggtccccccctcagctc 1736
Db 375 CAGCACTGGAATTTGCTTACAGGAATCTGGCTACAGCGACTGGGTCCCCCTCAGTCT 316
Qy 1737 ttcagcaacctggaccgaaacccggtggtgcacatgaacggatcacctcgtcaactgcaaaa 1796
Db 315 TTCAGCAACCTGGGACCGCAACCGGTGGCAATGAACGATCACCTTGTCAACTGGCAAA 256
Qy 1797 agtctatgcttttcaaacatcacagccagcccatatgagaggggtgagacctatctaaacgg 1856
Db 255 AGTGTATGGCTTTTCAACATCACAGCCGACCATATGAGAGGGTGGACCTATCTTAACAGG 196
Qy 1857 tatccaggaatcgtagaagaagctcctacgaggtcctcacagttccaacaaactcagtg 1916
Db 195 TATCCAGGAATCGTGAAGAAGCTCTACGGAGGCTCTCACAGTTCAACAAACTGCAGTG 136
Qy 1917 ccggtcaggtatcccccaagaccccaagagtaaaccttagtctcaatggagggtctctag 1976
Db 135 CCGGTCAAGTATCCCCCAAGACCCCAAGAGTAACCCCTAGGCTCAATGGAGGGTCTGG 76
Qy 1977 ggaccatggtatagagagaaacccaagaaaaagaaagcaagcaaaaaatcaggctgagaaa 2036
Db 75 GGACCATGTGTATAAGAGAGAAACCAAGAAAAAGAACCCAAAGCAAAAATCAGGCTGAGAAA 16
Qy 2037 aagcaaaaagaaaagc 2051
Db 15 AAGCAAAAAGAAAAAGC 1

RESULT 4
AZ815176
LOCUS
DEFINITION 2M0083G13F Mouse 10kb plasmid UUC1M library Mus musculus genomic
clone UUC2M0083G13 F, DNA sequence.
ACCESSION AZ815176
VERSION AZ815176.1 GI:12985084
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 625)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.,
and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0083 row: G column: 13
Seq primer: CGTTGTAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 625.
FEATURES
source
1..625
/organism="Mus musculus"
/strain="C57Bl/6J"

```

```
/db_xref="taxon:10090"
/clone="UUC2M008G13"
/clone_lib="Mouse 10kb plasmid UUC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gii147321141gbiAF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      179 a  150 c  159 g  137 t
ORIGIN

Query Match      22.6%; Score 512.4; DB 12; Length 625;
Best Local Similarity 90.0%; Pred. NO. 1.8e-88;
Matches 549; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 1237 acaacgtgacattgctctaaagacttatggtttctataaacacagcattatcttact 1296
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 16 ACAATGTGACCCCTGAAGAGGTATGGTTCTTAACAATAGCATTATCAITTTACT 75

QY 1297 ctccagataatgtgtgccagctcagcgagggaggaagtaactggcctctcagaggtagca 1356
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 76 CCTCAGATAATGTTGGCGAGCCACAGCAGGAGGAAGTAACCTGGCCACTCAGAGGCAGCA 135

QY 1357 aaggaacatattggaaaggagatcccggtgtgtaggctttgtgcatagcccaactctga 1416
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 136 AAGGAACATATTGGGAAGGAGGCATCCGGGCACTTGGCTTTGTGCATAGCCCACTCTAA 195

QY 1417 aaaaacagggaacagtgtgtaaggaaacctgtgcacactgactgactgctacccactctca 1476
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 196 AAAACAGGGAACGATGTGTAGGAACCTGTGCACATCACCATTGGTACCCCACTTGA 255

QY 1477 ttctactggctgaaggacagattgatgaggacattcaactagatggctatgatctgagg 1536
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 256 TTTCACTGGCTGAAGCAGACAGATTGATGAAGACATTTCAGCTAGATGATCGATATCTGGG 315

QY 1537 agacataaagtgggtttctgctcaaccccgagtagatattttgataaacattgacccca 1596
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 316 AGACCATAAGCGAAGGCTTCTGTCACCCGAGTGGATATTTTGACACAACATTTGACCCCA 375

QY 1597 tataccaaaggcaaaaatgctctctggcgagcgaggtatgggtatgggaactgcaaa 1656
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 376 TTTACACCAAGCGGCAAAATGGCTCTCTGGCGAGCAGGCTATGGGATCTGGAACACATGCA 435

QY 1657 tcacgtcagccatcagatgcagcactggaattgtcttacaggaataccttggtacagcg 1716
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 436 TCCAGCTGGCCATCGGGGTACAAACACTGGAACACTGGAACACTGCTCAGGCAATCCTGGCTACAGTG 495

QY 1717 actgggtcccccctcagctcttttcagcaacctgggacccagcgggtggcacaatgaacgga 1776
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 496 ACTGGGTCCCCCTCAGGCTTTACAGATCTCTGGGCCCCAAACCGGTGGCACAATGAAGAGA 555

QY 1777 tcacctgtcaactggcaaaagtgtatggcttttcaacatacacagccgacccatagtaga 1836
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 556 TTACCTTGTCAACTGCGCAAGAGTATCTGGCTTTTCAACATCAGCGCTGATCCGTATGAGA 615

QY 1837 ggggtggaacct 1846
```

```
|||||
Db 616 GGGTGGCACT 625

RESULT 5
BB520951
LOCUS BB520951
DEFINITION BB520951 RIKEN full-length enriched, 16 days neonate heart Mus
          musculus CDNA clone D830047F08 3', mRNA sequence.
ACCESSION BB520951
VERSION BB520951.2 GI:16443460
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 648)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
          Hiramoto,K., Hori,F., Ishi,Y., Ito,M., Kawai,J., Konno,H., Kouda
          ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
          Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
          ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
          Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
          Muramatsu,M. and Hayashizaki,Y.
          RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
          Unpublished (2001)
          On Jul 28, 2000 this sequence version replaced gi:9572409.
          Contact: Yoshihide Hayashizaki
          Laboratory for Genome Exploration Research Group, RIKEN Genomic
          Sciences Center(GSC), Yokohama Institute
          The Institute of Physical and Chemical Research (RIKEN)
          1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
          Tel: 81-45-503-9222
          Fax: 81-45-503-9216
          Email: genome-res@gsc.riken.go.jp,
          URL:http://genome.gsc.riken.go.jp/
          Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
          ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
          Normalization and subtraction of cap-trapper-selected cDNAs to
          prepare full-length cDNA libraries for rapid discovery of new
          genes. Genome Res. 10 (10), 1617-1630 (2000)
          wagi,K., Fujiwaki,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
          Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
          ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
          Hayashizaki,Y.
          RIKEN integrated sequence analysis (RISA) system--384-format
          sequencing pipeline with 384 multicapillary sequencer. Genome Res.
          10 (11), 1757-1771 (2000)
          Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
          ,Y. and Hayashizaki,Y.
          Computer-based methods for the mouse full-length cDNA
          encyclopedia: real-time sequence clustering for construction of a
          nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
          Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamana,K.I., Aizawa
          ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
          Hayashizaki,Y.
          Computational Analysis of Full-Length Mouse cDNAs Compared with
          Human Genome Sequences Mamm. Genome 12, 673-677 (2001)
          Please visit our web site (http://genome.gsc.riken.go.jp/) for
          further details.
          cDNA library was prepared and sequenced in Mouse Genome
          Encyclopedia Project of Genome Exploration Research Group in Riken
          Genomic Sciences Center and Genome Science Laboratory in RIKEN.
          Division of Experimental Animal Research in Riken contributed to
          prepare mouse tissues.
          Location/Qualifiers
            1..648
              /organism="Mus musculus"
              /db_xref="taxon:10090"
              /clone="D830047F08"
              /clone_lib="RIKEN full-length enriched, 16 days neonate
              heart"
              /tissue_type="heart"
```

```
/dev_stage="16 days neonate"
/lab_host="DH10B"
/notes="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GACAGAGAGCGCGCACTCGAGTGTGTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GACAGAGAGATTGTCGAGTGAATTAATTAATCCCGCCCCCCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC I."
```

BASE COUNT	179 a	163 c	171 g	135 t
ORIGIN				
Query Match	22.6%	Score 511.2;	DB 9;	Length 648;
Best Local Similarity	87.8%;	Pred. No. 3.1e-88;		
Matches 569;	Conservative 0;	Mismatches 78;	Indels 1;	Gaps 1;

```
QY 1346 cagaggtagcaaaaggaacatatattgggaaggagggatccgggctgtagcttggcatag 1405
      ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 1 CAGAGGCGAGCTAAGGACAAATATTGGAAGAGCATCCGGCGACTGTGCTTGTGCATAG 60

QY 1406 ccaactctgaaaaaaa-gggaacagtgtgtaagggaacctgtgcatactactactggt 1464
      ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 61 CCCACTTCTTAAAGAACCAAGGGGACGGTATGTGAAGAACTTGTGCATCATCCGATTGGT 120

QY 1465 acccaactctcttactggtgaagacagattgatgagacattcaactagatggct 1524
      ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 121 ACCCACTTGTATTCTACTGCTGAAGCAGACAGATTGATGAACATTCAGCTAATGGAT 180

QY 1525 atgatctggagaccataagtggaggtctctgctcaccccgagtatatttggcata 1584
      ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 181 ACGATATCTGGAGACCATTAAGCAAGGCTCTTCGTTACCCCGAGTGATATTTGCACA 240

QY 1585 acattgaccccatatacacaagcaaaaaatgctctctggtggcagcgctatgggatct 1644
      ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 241 ACATTGACCCCATTTACACCNAAGGCGAAAAATGGCTCTCGGCGACGAGCTATGGGATCT 300

QY 1645 ggaacactgcaatccagtcaagccatcagagtgcagcactggaattgtttacaggaatac 1704
      ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 301 GGAACACTGCAATCCAGTCGGCCATCCGGGTACAACTGGAACACTGCTCACAGGCAATC 360

QY 1705 ctggctacagcagtggtgtcccccctcagctctttcagaacctgggaccgaaacgggtggc 1764
      ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 361 CTGGCTACAGTGACTGGGTCCCCCTCAGGCTTTCAGCAATCTGGGCCCAACCCGGTGGC 420

QY 1765 acaatgaacgatcacctctgcaactggcaaaagtgtatgagctttcaacatcacagccg 1824
      ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 421 ACAATGAAGGATTACCTGTGCAACTGCAAGAGATATCTGCGCTTTTCAACATCACGGCTG 480

QY 1825 acccatatgagaggttgagacctatctaacaggttatccaggaatcgtgaagagctctac 1884
      ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 481 ATCCGTATGAGAGGGTGGACCTATCCACAGCGTATCCAGGCGATCGTGAAGAGCTGTGCG 540

QY 1885 ggaaggtctctcagttcaacaaaaactcagtgccggtgagtgatctcccccccaaaacccca 1944
      ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 541 GGAGACTGTCAATTAACAAGAGCTGCGGTGCTGTGAGGTACCCCGGAGGATCCCA 600

QY 1945 gaagtaaccttagctcaatggaaggtctctaggaacctatggtatag 1992
      ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 601 GAAGCAACCTTAGCTCAACGGAGAGTCTGGGGTCCGTGCTGTATAAAG 648
```

RESULT 6

BB319509 BB319509 700 bp mRNA linear EST 24-OCT-2001  
LOCUS

DEFINITION  
BB319509 RIKEN full-length enriched, adult male corpora  
quadrigenina Mus musculus cDNA clone B230380L03 3', mRNA sequence.  
BB319509  
VERSION  
BB319509.2 GI:16403267  
KEYWORDS  
EST.  
SOURCE  
house mouse.  
ORGANISM  
Mus musculus

REFERENCE  
AUTHORS

TITLE  
JOURNAL  
COMMENT

Unpublished (2001)  
On Jul 11, 2000 this sequence version replaced gi:9026544.  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.go.jp  
URL: http://genome.gsc.riken.go.jp/  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh  
.M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
wagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura  
.S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and  
Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicepillary sequencer. Genome Res. 10  
(11), 1757-1771 (2000)  
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara  
.Y. and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Alizawa  
.K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and  
Hayashizaki, Y.  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Please visit our web site (http://genome.gsc.riken.go.jp/) for  
further details.

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.

FEATURES  
source

Location/Qualifiers  
1..700  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="B230380L03"  
/clone\_lib="RIKEN full-length enriched, adult male corpora  
quadrigenina"  
/sex="male"  
/tissue\_type="corpora quadrigenina"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Site\_1: Sali; Site\_2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in



Qy	1625	ggcagcagcctatggatctggaacactgcaatccacgctcagcgcacatcaagatgcagcactg	1684
Db	413	GCACAGGCGATGGGATCTGGAACTCTGAATCCAGTCAGGCATCAGCATGTCAGCACTG	354
Qy	1685	gaatgtctcacaggaaatcctggctacagcagctgggtccccctcagttcttttcagcaa	1744
Db	353	GAAATGCTTTACAGGAAATCTCGCTACAGCACT - GGTCCCCCTCAGTCTTTTCAGCAA	295
Qy	1745	cctgggaccgaaacccgtgtggcacaatgaaacggtacacctgtcaactggcacaagtgatg	1804
Db	294	CTTGGACCAACCGGTGGGCACAATCAACCGATCACCTTGTCACTGGCAAAAGTGTATG	235

[illegible]

Db	Accession	Sequence	Length
Db 174	AATCGTGAAG -	ACAGTCTACGGAGGCTCTCACAGTTCACAAAACTGCAGTGCCGGTCAG	116
Qy 1925	gtatccccccaaagaccccaagaagtaacccctaggctcaatgagagggtctaggagaccatg	1984	
Db 115	GTATCCCCCAAGACCC	CAGAAGTAACCTTAGGCTCAATGGA - GGGTCTGGGGACCATG	57
Qy 1985	gtatagagaggaaccaaagaaaaagcccaagcaaaaatcaggctgagaaaaagc	2040	
Db 56	GTATAAGAGGAAACCAAGAAAAAAGCAAGCAAAAATCAGGCTGAGAAAAAGC	1	

[illegible]

REFERENCE  
AUTHORS  
1 (bases 1 to 459)  
Smith, T. P. L., Grosse, W. M., Freking, B. A., Roberts, A. J., Stone, R., Casas, E., Wray, J. E., White, J., Cho, J., Fahrenkrug, S. C., Bennett, G. L., Heaton, M. P., Laenreid, W. W., Rohrer, G. A., Chittock-McKown, C.

**TITLE**  
 Journal of Molecular Evolution  
**JOURNAL**  
 MEDLINE  
**COMMENT**  
 Pertea, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J., and Keefe, J. W.  
 Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle  
 Genome Res. 11 (4), 626-630 (2001)  
 21180013  
 Contact: Smith TPL  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smith@email.marc.usda.gov  
 Single pass sequencing. Bases called and alt\_trimmed with phred v0.980904.e. Vector identified by cross\_match with the -minscore and -minmatch 12 options.  
 PCR primers  
 FORWARD: AGGAACACGCTATGACCAT  
 BACKWARD: GTTTCACGTCACGACG  
 Plate: 66 row: F column: 20

```
FEATURES
source
seq primer: ATTAGTGACACTATAG..
Location/Qualifiers
1..459
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH10B"
```

Library made from pooled tissue from marrow, alveolar macrophage, ovary, fetal semitendinosus muscle, and fetal longissimus muscle."

BASE COUNT	118 a	118 c	119 g	104 t		
ORIGIN						
Query Match	15.9%; Score 361.4; DB 10; Length 459;					
Best Local Similarity	86.7%; Pred. No. 1.9e-59;					
Matches	398; Conservative	0; Mismatches	61; Indels	0; Gaps	0;	
QY	1426	gaacagtgtgaaggaaacctgtgcacatcactgactggtgtacccactctcttcactgg	1485			
Db	1	GCACCGTGTGAAGGAGCTGTGCACATCACTGCTGTACCCACTCTGATTTCACTGG	60			
QY	1486	ctgaaggacagatgatgagagacattcaactagatgcttatgatctctggagagaccataa	1545			
Db	61	CTGAGGGGCAGATTGATGAGAACATTCACATGGATGGCTATGATGCTCTGGGAGACCATAA	120			
QY	1546	gtgaggggtcttcgctccaccgagtagatatttgcataacattgacccatatacaccca	1605			
Db	121	CGAAGGCCCTTCGTCTCCCGTGTAGACATTCGCACAACATTGACCCATTATACCA	180			
QY	1606	agggcaaaaatgctcctggcagcaggtgatggatctggaacactgcaatccagtcag	1665			
Db	181	AGGCGAAAAATGGCTCTTGGGCTGCAGGCTATGGATCTGGAACACTGCCATTCACTGTG	240			
QY	1666	ccatcagagtcgagcactggaattgcttacaggaatactcctggctacagcactgggtcc	1725			
Db	241	CTATCAGGGTGAACACTGGAAATTCACACAGGGAACCCAGGATATAGTACTGGGTGC	300			
QY	1726	ccctcagctcttcagcaacctgggacgcagcgggtggcacaatgaacggatcacctcgt	1785			
Db	301	CCCTCTAGTCTTTCAGCAACTTGGGGCAACACCGGTGGCACAATGAGCGAATTACCTTGT	360			
QY	1786	caactggcaaaagtgtatggcttttcaacatcacagccgaccatattagaggggtggacc	1845			
Db	361	CAACGGGAAAAAGCGTGTGGCTGTTCACATTACAGCGCGACCCCTATGAGAGAGTAGACC	420			
QY	1846	tatctaacaggtatccaggaaatcgtgaagaagctcctac	1884			
Db	421	TTTCAACAGGATATCCCGGGATCGTAAAGCAGCTCCTCC	459			
RESULT	10					
LOCUS	BG661302	547 bp	mrna	linear	EST 05-JUL-2001	
DEFINITION	la59e07.y1 Melton Human Islets H1Z1 Homo sapiens cDNA 5', mRNA					
ACCESSION	BG661302					
VERSION	BG661302.1	GI:13816937				
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	1 (bases 1 to 547) Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K., Lemishka, I., Scarce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R., Williams, T., Jackson, Y. and Bowers, Y.					
TITLE	Endocrine Pancreas Consortium					
JOURNAL	Unpublished (2000)					
COMMENT	Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue Endocrine Pancreas Consortium Harvard University, Howard Hughes Medical Institute Dept. of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138 Tel: 617-495-1812 Fax: 617-495-8557 Email: dmelton@biohph.harvard.edu					

Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Juliana Brown (brownjefas.harvard.edu)  
Seq primer: -40RP from Gibco  
High quality sequence stop: 482.

FEATURES		source		high quality sequences (sup: 40% Location/Qualifiers).	
		1. .547			
		/organism="Homo sapiens"			
		/db_xref="taxon:9606"			
		/clone_lib="Melton Human Islets H1Z1"			
		/sex="Both"			
		/tissue_type="Islets of Langerhans"			
		/dev_stage="Adult"			
		/lab_host="TOP10"			
		/note="Organ: Pancreas; Vector: p2BrO-2; Site_1: Not I; Site_2: Xho I; Library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Xho I site destroyed during cloning. Size-selected by column fractionation; average insert size 1.59 kb. Primary library, unamplified."			
BASE COUNT		169 a	124 c	120 g	134 t
ORIGIN					
Query Match		15.1%; Score 341.8; DB 10; Length 547;			
Best Local Similarity		99.2%; Pred. No. 1.1e-55;			
Matches		354; Conservative	0; Mismatches	2; Indels	1; Gaps 1;
QY	1897	agttcaacaaactgcagtcggtgcaggtatcccccaagaccccgagaagtaacccta	1956		
Db	1	AGTTCAACAAACTGCAGTCGCGGTCAAGTATCCCCCAAGACCCAGAGTAACCCCTA	60		
QY	1957	ggctcaatggagggtctaggaccatggtatagagagaacccaagaaaagaagccaa	2016		
Db	61	GGCTCAATGGAGGGGTCTGGGGACCATGGTATAAAGAGGAAACCAAGAAAAGAGCCAA	120		
QY	2017	gcataaatcagcgtgagaaaaagcaaaagaaaagc-aaaaaagagaagaaacagcaga	2075		
Db	121	GCAAAAATCAGCGTGAAGAAAAGCAAAAGAAAAGCAAAAGAAAAGAGAAAGACGAGA	180		
QY	2076	aagcagtcctcaggttcaacttcaggttacttctggtgataagcacaataatttc	2135		
Db	181	AAGCAGTCTCAGGTTCAACTTGCATTCAGGTGTACTTGTGTAAGCACAATATTTC	240		
QY	2136	ctgtttggttaactttaatcagttcttattcttcttcattctggttcttcaggttaaacagcaa	2195		
Db	241	CTGTTTGTGTTAACTTTAATCAGTTCTTATCTTTTCATCTGTTTCTAGGTAAACAGCAA	300		
QY	2196	atttggtctgataatcgtcgtggcctaagcgtcaggtctgtttctgctgtgcccac	2252		
Db	301	ATTGGCTCGATATATCGCTGGCCTAAGCGTCAGGCTGTTTTCATGCTGTGCCAC	357		
RESULT 11					
LOCUS		W81485 344 bp mRNA linear EST 17-OCT-1996			
DEFINITION		2d87a04.r1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone IMAGE:347598 5' similar to SW:ARSB_HUMAN P15848 ARYL SULFATASE B PRECURSOR ;, mRNA sequence.			
ACCESSION		W81485			
VERSION		W81485.1 GI:1392515			
KEYWORDS		EST.			
SOURCE		human.			
ORGANISM		Homo sapiens			
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS		Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.			
TITLE		The WashU-Merck EST Project			

JOURNAL  
COMMENT

Unpublished (1995)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu

This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 546 Std Error: 0.00

Seq primer: mob.REGA+ET

High quality sequence stop: 303.

FEATURES  
source

Location/Qualifiers  
1. .344  
/organism="Homo sapiens"  
/db\_xref="GDB:1272973"  
/db\_xref="taxon:9606"  
/clone="IMAGE:347598"  
/clone\_lib="Soares\_fetal\_heart\_NbHH19W"  
/sex="unknown"  
/dev\_stage="19 weeks"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: heart; Vector: pT7m3D (Pharmacia) with a  
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACCAATCTGAAGTGGGCGCGCATCTTTTTTTTTTTTTTTT 3']  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by  
M.Fatima Bonaldo. This library was constructed from the  
same fetus as the fetal lung library, Soares fetal lung  
NbHL19W."

BASE COUNT 92 a 93 c 82 g 74 t 3 others

ORIGIN

Query Match 14.7%; Score 333.2; DB 10; Length 344;  
Best Local Similarity 98.2%; Pred. No. 5.5e-54;  
Matches 335; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1551 ggcttcgctcaacccagtagatatttgcataaacattgaccccatatacacaagca 1610  
|||||  
Db 1 GGTCTTCGCTCACC CGAGTAGATATTTCATACATTGACCCCATATACCAAGCA 60  
|||||

QY 1611 aaaaaggctctggcagcagctatggatctggaacctgcaatccactcagccatc 1670  
|||||  
Db 61 AAAAATGGCTCCTGGGCAGCAGGCTATGGGATCTGGAAACATGCAATCCATCAGCCATC 120  
|||||

QY 1671 agagtgcagcactggaattgcttacaggaaatcctggctacagcagctgggtcccccct 1730  
|||||  
Db 121 AGAGTGCGAGCACTGGAAATGCTTACAGGAAATCCTGGCTACAGCGACTGGTCCCCCT 180  
|||||

QY 1731 cagttttcagcaacctgggaccgaacgggtggcacaaatgaacggatcacctcgtcaact 1790  
|||||  
Db 181 CAGTCCTTTTCAGCAACCTGGGACCGAACCNGTGGCACATATGAACGATCACCTTGTCAACT 240  
|||||

QY 1791 ggcaaaagtgtatggcttttcaacatcacagccagccaccatgatgagggtagacctatct 1850  
|||||  
Db 241 GGCRAAAGTGATGGCTTTTTCACATCATCACGCCGACCCCATATGAGAGGTGGACCTATCT 300  
|||||

QY 1851 aacagggtatccaggaatcgtgaagaagctctctacggagct 1891  
|||||

Db 301 AACAGGTATCCATGANTCGTGAAGANGCTCCTACGGGGCT 341  
|||||

RESULT 12

R43179/c

LOCUS

DEFINITION R43179 458 bp mRNA linear EST 22-MAY-1995  
Y018403.s1 Soares infant brain INIB Homo sapiens cDNA clone  
IMAGE:32854 3' similar to SP:ARSB\_HUMAN P15848 ARYL SULFATASE B  
PRECURSOR ; , mRNA sequence.

ACCESSION  
VERSION

R43179.1 GI:825405

KEYWORDS  
SOURCE

EST.  
human.

ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 458)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman  
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,  
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston  
R., Williamson, A., Wohlmann, P. and Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)

TITLE

JOURNAL

COMMENT

On May 8, 1995 this sequence version replaced gi:801403.

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

Insert Size: 1275

High quality sequence stops: 384 Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1275 Std Error: 0.00

Seq primer: Promega -21ml3

High quality sequence stop: 384.

FEATURES

source

Location/Qualifiers  
1. .458  
/organism="Homo sapiens"  
/db\_xref="GDB:405201"  
/db\_xref="taxon:9606"  
/clone="IMAGE:32854"  
/clone\_lib="Soares infant brain INIB"  
/sex="female"  
/dev\_stage="73 days post natal"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: whole brain; Vector: Lfamid BA; Site.1: Not  
I; Site.2: Hind III; 1st strand cDNA was primed with a Not  
I - oligo(dT) primer [5'  
AAGTGAAGAAATCGCCGCCAGGAATTTTTTTTTTTTTTTT 3'];  
double-stranded cDNA was ligated to Hind III adaptors  
(Pharmacia), digested with Not I and directionally cloned  
into the Not I and Hind III sites of the Lfamid BA vector.  
Library went through one round of normalization. Library  
constructed by Bento Soares and M.Fatima Bonaldo."

74 a 110 c 119 g 152 t 3 others

BASE COUNT

ORIGIN

Query Match 14.6%; Score 331.2; DB 10; Length 458;  
Best Local Similarity 92.6%; Pred. No. 1.2e-53;  
Matches 401; Conservative 0; Mismatches 26; Indels 6; Gaps 5;

QY 1632 ggctatggatctggaacactgcaatcc--agtacgcatcagagtg-cagcactggaaa 1688  
|||||

Db 434 GCCTATGGGATCTGGAACTGCAATCCCATCCAGTCAGCCCATCAGATGCCAGCCTGGAAA 375  
|||||

QY 1689 ttgttacagaaat-cctggctacagcg-actgggtccccctcagttcttcagcaa-c 1745  
|||||

Db 374 TTNCTTACAGAAATCCCTGGCTACAGGAACTGGGTCCCCCTCAGTCTTTTCAGCAACC 315  
|||||

QY 1746 ctgggaccgaaccggtggcacaatgaacgcatcactcgtcaactggcaaaagtatagg 1805  
|||||

Db 314 CTGGACCGAACCGGTGGCACAATGAACGGATCACCTTGTCAACTGGCAAAAGTATGG 255  
|||||

QY 1806 cttttcaacatcacagccgacccatgatgagagggtggacctatctcaacaggtatccagga 1865  
|||||

Db 254 CTTTTCACATCAGACCGGACCCATATGAGAGGTGGACCTATCTTAACAGGTATCCCAGG 195  
|||||

QY 1866 atcgtgaagaagctcctcagcagggctctcagagttcaacaaactgcagtcgcggtcagg 1925  
|||||







```
Qy 2016 agcaaaatcaggctgagaaagcaaaagaaagcaaaaa-----aaagaagaagaaacag 2071
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 320 AACAAACCAAGGCTAAGAAATGCAAAAGAAAGAAAGTAAGGCAAGGATGAGGAAGCAG 379
Qy 2072 cagaagaagctcaggttcaacttgccatcaggtgtaacttggtaagcacaata 2131
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 380 TTGGCAGCTCACTCCTCAATAAATGCCACCCCAAGTGTGCCACTGGCTGAATAATGCAAAAC 439
Qy 2132 ttctcgtttggttaaaccttaacagttcttattcttcttcacgtctgttctcaggttaaacca 2191
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 440 TTTTCGCCCTGGTTAAACTTAGA-----CTGCCGTGGTTCTAGGTAATTCAC 485
Qy 2192 goaaatttggtcgataataatcgctggcctaagcgctcaggctgtttttcactgctg 2246
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 486 AAGAAATTTGGCTCAGAAATACCTCGGCCAACCGCTCGGGCTGTGTTTCACATTG 540

RESULT 15
BE553565 628 bp mRNA linear EST 15-AUG-2000
LOCUS ur47a08.y1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:3153398 5'
DEFINITION similar to SW:ARSB_FELCA P33727 ARYLSULFATASE B PRECURSOR ;, mRNA
sequence.
ACCESSION BE553565
VERSION BE553565.1 GI:9817982
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 628)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
This clone is available royalty-free through LInL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1056154
Seq primer: -40RP from Gibco
High quality sequence stop: 390.
FEATURES
source
1..628
/organism="Mus musculus"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone_lib="IMAGE:3153398"
/tissue_type="NCI_CGAP_Mam2"
/dev_stage="tumor, biopsy sample"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
BASE COUNT 141 a 201 c 165 g 121 t
ORIGIN
```

```
Query Match 11.6%; Score 262; DB 10; Length 628;
Best Local Similarity 64.6%; Pred. No. 2.1e-40;
Matches 407; Conservative 0; Mismatches 220; Indels 3; Gaps 1;

Qy 596 gggttaccaggtatcgagattaaacacctactcttgacagctcgctgcgaaggaggt 655
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 GGGATACCATGGCTCGGATATCGAGACCCCAACCGTGGACCGGCTAGCAGCTGAGGGGCT 60

Qy 656 taaactggagaaactatgtcagctatttgcacacattccacaggtcagttttattac 715
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 CAAACTGGAGAAATATTATACATCCAAACCCCATATGTATACACTTCGAGGAGTCAACTCCTCAC 120
```

```
Qy 716 tggaaagtatcagatacacacacgggacttcacattctatcataagacacacccca 775
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 CGGCAGGTACACAGATCCATACAGGATTGCAGCACTCCATTATCCGCCACGCGCAGCCCAA 180
Qy 776 ctgtttacctctggaacaatgcccacctacacctcagaataaaggagggttgatatcaac 835
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 CTGTCGTGCCCTGGACCAAGGTACGCTGCCCAAGAGCTACAGGAGGCGAGGCTACTCCAC 240
Qy 836 gcatagtcggaataatggcacttgggtttttacagaaaaaagatgcatgcccaccagaag 895
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 CCACATGTTGGCAAGTGGCATCTCGGGCTTCTACCGGAAGGAGTGTTCCTTACCTACCGCGCG 300
Qy 896 aggtattgataacctttttgggtcccttttgggaagtggggtattactatcacacactaca 955
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 301 GGGCTTCGACACACTTCCTGGGTTCCTCACAGGCAATGTGGATTACTACACCTACGACAA 360
Qy 956 atgtgaagtcctgggatgtgtggctatgacttgtatgaaaacgacaatgctgcctggga 1015
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 CTGTGATGGCCCGAGGGGTTGTGGGTTTGACCTGCACGAGGTGAGAGGCTGGCCTGTG - 419
Qy 1016 ctatgacaatggcatatatactccacacagatgtacactcagagagtacagcaaatcttagc 1075
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 420 --GGCTCAGCGGCCAGTACTCCACTATGCTCTACGCTCAGCGTGCAGGCCACATCTGCG 477
Qy 1076 ttcccataacccccacaaagcctatatattttatatatttgcctatcaagctgttctacc 1135
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 478 CAGGCACAAATCCCCAGAAATCCTCTCTTCTCTATGTGGCTTCCAGGAGGTACACACGCC 537
Qy 1136 actgaaactcctgcaggattattcgaacactacccatccattatcaacataaacaggag 1195
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 538 CCTACAGTCACTTCGAGAAATACCTATACCGCTTACCGCAGATGGGCAACGTTAGCAGCG 597
Qy 1196 gagatatgctgccatgcttctcgtcttaga 1225
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 598 CAAGTACCGCAGCATGGTGACCTGCATGGA 627
```

Search completed: July 23, 2002, 15:52:09  
Job time: 3130 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 15:08:04 ; Search time 2392.27 Seconds  
(without alignments)  
12784.552 Million cell updates/sec

Title: US-09-495-823-8  
Perfect score: 2266  
Sequence: 1 cacggtccgcacgcgc.....tgccacctgtgcgaattc 2266

Scoring table: OLIGO\_NUC

Gapex 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

- EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	505	22.3	817	10	BG680752 602628445
C 2	465	20.5	615	9	AW474222 XS20N05.x
3	382	16.9	683	10	BF307503 601893984
C 4	212	9.4	735	10	BG772761 602720927
5	210	9.3	344	10	W81485 zdB7a04.r1
6	201	8.9	547	10	BG661302 Ia59e07.y
C 7	200	8.8	1032	10	BF979627 602287919
8	155	6.8	274	10	BG194726 RST13852
9	119	5.3	552	10	BF218442 601881811
10	115	5.1	325	10	U46280 HSU46280.Hu
C 11	97	4.3	414	10	W81455 zdB7a04.s1
C 12	94	4.1	458	10	R43179 yG18d03.s1
C 13	94	4.1	463	9	AA159953 zO75a09.s
14	81	3.6	849	10	BF696654 602124758
15	70	3.1	407	9	AA159952 zO75a09.r
16	55	2.4	951	10	BG167651 602342825
17	53	2.3	625	12	AZ815176 2M0D83G13

18	53	2.3	648	9	BB520951
19	53	2.3	700	9	BB319509
20	40	1.8	265	9	BE149009
21	33	1.5	61	10	R20122
22	30	1.3	459	10	BF653682
23	29	1.3	388	9	BE047859
24	28	1.2	717	10	BG975362
25	27	1.2	338	10	B1693994
26	27	1.2	397	10	BG899940
27	27	1.2	552	9	AL022653
28	27	1.2	575	9	AV596948
29	27	1.2	625	10	BF537953
30	27	1.2	692	10	BI328079
31	27	1.2	724	10	BI697980
32	26	1.1	185	9	AW429593
33	26	1.1	277	9	AV615802
34	26	1.1	337	10	BM495947
35	26	1.1	479	9	AI641294
36	26	1.1	506	9	AI497032
37	26	1.1	514	9	AI877708
38	26	1.1	543	9	AV617615
c	39	26	565	10	BM280523
40	26	1.1	602	9	AV595356
41	26	1.1	604	9	AV595699
42	26	1.1	703	10	BI109023
43	26	1.1	774	10	BI689140
44	26	1.1	804	9	AI545901
45	26	1.1	806	10	BI684551

ALIGNMENTS

RESULT 1

BG680752

LOCUS 602628445F1 NCI\_CGAP\_Skn4 Homo sapiens cDNA clone IMAGE:4753259 5', mRNA linear EST 01-MAY-2001  
DEFINITION mRNA sequence.

ACCESSION BG680752

VERSION BG680752.1 GI:13912149

KEYWORDS EST..

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 817)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: James Cleaver, M.D.

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM0612 row: j column: 12

High quality sequence stop: 789.

Location/Qualifiers

1..817

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4753259"

/clone\_lib="NCI\_CGAP\_Skn4"

/tissue\_type="squamous cell carcinoma"

/lab\_host="DH10B (T1 phage-resistant)"

/note="Organ: skin; Vector: pCMV-SPORT6; Site: 1; Not I;

Site: 2; Sall; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.5Kb. Library constructed by Life

Technologies. Note: this is a NCI\_CGAP Library."

BASE COUNT 259 a 193 c 207 g 158 t

ORIGIN

Query Match 22.3%; Score 505; DB 10; Length 817;  
Best Local Similarity 99.7%; Pred. No. 3.1e-197;  
Matches 605; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1368 tgggaaggagggatcgccgctgtagcttctgtcagtagccactctctgaaacaaggga 1427  
|||||  
Db 9 TGGGAAGAGGAGATCGGGCTGTAGCTTGTGTCATAGCCCACTTCTGAAAAACAGGGA 68  
|||||

Qy 1428 acagtgtgaaggaacctgtgcacatcactgactggtgaccccaactctctatttcaactggct 1487  
|||||  
Db 69 ACAGTGTGTGAAGCAACTTGTGCACATCACTGACTGCTGTAACCCACCTCTCATTTCACTGGCT 128  
|||||

Qy 1488 gaagacagattgatgagacaattcaactagatgctctgatatctgtgagagaccataagt 1547  
|||||  
Db 129 GAAGGACAGATTGATGAGACATTCAACTAGATGGCTATGATATCTGGGAGACCAATAAGT 188  
|||||

Qy 1548 gaggtcttcgctcaccgagtagatatattgcatataacattgaccccatatcaccaag 1607  
|||||  
Db 189 GAGGTCTTCGCTCACCAGAGTAGATATTTTGCAATAACATTGACCCCATATACACCAAG 248  
|||||

Qy 1608 gcaaaaaatggtctcctggtggcagcaggtctatggtatctggaacactgcaatccagtcagcc 1667  
|||||  
Db 249 GCAAAAAATGGCTCCTGGGCAGCAGGCTATGGGATCTGGAACACTGCAATCCAGTCAGCC 308  
|||||

Qy 1668 atcagagtgcagcactggaattgcttacaggaatactctggtctacagcagactgggtccccc 1727  
|||||  
Db 309 ATCAGAGTGCAGACTGGAATTTGCTTTACAGGAATCTCTGGCTACAGCGACTGGGTCCCC 368  
|||||

Qy 1728 cctcagctcttcagcaacctgggacccgagaccggtggcacaatgaaacggtacacctcgctca 1787  
|||||  
Db 369 CCTCAGTCTTTTCAGCAACTGGGACCGACCGTGGCACAATGAACGGATCACTTGCTCA 428  
|||||

Qy 1788 actgccaagtgatggtcttcaacatcacagccagcccatatgagaggtgagacctta 1847  
|||||  
Db 429 ACTGGCAAAAGTGTATGGCTTTTCAACATCACAGCGCAGCCCATATGAGAGGTTGGACCTA 488  
|||||

Qy 1848 tctaacaggtatccaggaactggaagctcctcagcaggtctcagaggtctcagagttcaacaaa 1907  
|||||  
Db 489 TCTAACAGGTATCCAGGAATCTGGAAGAGCTCCTACGAGGAGCTCTCACAGTTTCAACAAA 548  
|||||

Qy 1908 actgcagtcggtcaggtatcccccccaagaccccgagaaagtaacccctaggtcgaatgga 1967  
|||||  
Db 549 ACTGCAGTCCGGTCAGGTATCCCCCAGAGACCCAGAGATACCCCTAGGCTCAATGGA 608  
|||||

Qy 1968 ggggtct 1974  
|||||  
Db 609 GGGGTCT 615  
|||||

RESULT 2  
AW474222/c  
LOCUS  
DEFINITION  
xs20h05.x1 NCI\_CGAP\_Ut2 Homo sapiens cDNA clone IMAGE:2770233 3'  
similar to SW:ARSB\_HUMAN PI5848 ARYL SULFATASE B PRECURSOR ;, mRNA  
sequence.

ACCESSION  
AW474222  
VERSION  
AW474222.1  
KEYWORDS  
GI:7044328  
SOURCE  
human.

ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
1 (bases 1 to 615)  
AUTHORS  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL  
Unpublished (1997)  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution Information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/dbp/image/image.html](http://www-bio.llnl.gov/dbp/image/image.html)  
Seq primer: -400p from Gibco  
High quality sequence stop: 435.  
Location/Qualifiers  
1. .615  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2770233"  
/clone\_lib="NCI\_CGAP\_Ut2"  
/tissue\_type="moderately-differentiated endometrial  
adenocarcinoma, 3 pooled tumors"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: Salt;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.85 kb. Life Technologies catalog #:  
11539-012"

BASE COUNT 116 a 151 c 153 g 195 t  
ORIGIN

Query Match 20.5%; Score 465; DB 9; Length 615;  
Best Local Similarity 99.8%; Pred. No. 1e-180;  
Matches 515; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1459 actggtaccacctctcattcactggtgagagcagattgatgaggacattcaactag 1518  
|||||  
Db 593 ACTGGTACCCCACTCTCATTTTTCACCTGGCTGAAGGACAGATTGATGAGGACATTCAACTAG 534  
|||||

Qy 1519 atgctatgatctctggtgagaccataagtgggtcttcgctcaccctcagtagatatt 1578  
|||||  
Db 533 ATGGGTATGATATCTGGGAGACCAATAAGTGAAGGTCTTCGCTCACCCCGAGTAGATATT 474  
|||||

Qy 1579 tgataaacattgaccccatatcaccaagcacaataatgctcctggcagcaggtatg 1638  
|||||  
Db 473 TGCTAACATTGACCCCATATACCAAGGCAAAAATGGCTCTGGGACGAGGCTATG 414  
|||||

Qy 1639 ggatctgggaacctgcaatccagtcagccatcagagtcagcagcactggaatctgctacag 1698  
|||||  
Db 413 GGATCTGGAACACTGCAATCCAGTCAGCCATCAGAGTCAGCACCTGGAATTCCTTACAG 354  
|||||

Qy 1699 gaaatcctggtctacagcagctgggtcccccctcagctcttcagcgaacctgggacgaacc 1758  
|||||  
Db 353 GAAATCCTGGCTACAGCGACTGGGTCCGCCCTCAGTCTTTCAGCAACCTGGGACCGAACC 294  
|||||

Qy 1759 ggtggcacaatgaacggatcacctcgtcaactggcaaaagtgtatgctcttcaacatca 1818  
|||||  
Db 293 GGTGGCACAAATGACGGATCACCTTGTCAACTGGCAAAAAGTGTATGGCTTTTCAACATCA 234  
|||||

Qy 1819 cagccgaccatgatgagaggtggtgacctatctaaacaggtatccaggaatcgtgaaagaagc 1878  
|||||  
Db 233 CAGCCGACCCATATGAGAGGGTGGACCTATCTAACAGGTATCCAGGAATCGTGAAGAAGC 174  
|||||

Qy 1879 tctacggaggtctctcaagttcaacaacaaactgcagtcggtcaggtatcccccaag 1938  
|||||  
Db 173 TCCTACGGAGGCTCTCACAGTTTCAACAAAACCTGCAGTGGCGGTACAGGTATCCCCCAAG 114  
|||||

Qy 1939 acccgaagtaacctaggtcaatggaggggtct 1974  
|||||  
Db 113 ACCCCAGAAATACCCCTAGGCTCAATGGAGGGGTCT 78  
|||||

RESULT 3  
BF307503  
LOCUS  
DEFINITION  
BF307503  
601893984F1 NIH\_MGC\_17 Homo sapiens cDNA clone IMAGE:4139565 5',  
mRNA sequence.  
ACCESSION  
BF307503



IMAGE:347598 5' similar to SW:ARSB\_HUMAN P15848 ARYL SULFATASE B  
PRECURSOR ;, mRNA sequence.  
ACCESSION W81485  
VERSION W81485.1 GI:1392515  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 344)  
REFERENCE Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman  
AUTHORS M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,  
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston  
R., Williamson, A., Wohlmann, P. and Wilson, R.  
TITLE The WashU-Merck EST Project  
JOURNAL Unpublished (1995)  
COMMENT Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnlnl.gov) for further information.  
Insert Length: 546 Std Error: 0.00  
Seq primer: mob.REGA+ET  
High quality sequence stop: 303.  
FEATURES Location/Qualifiers  
1..344  
/organism="Homo sapiens"  
/db\_xref="GDB:1272973"  
/db\_xref="taxon:9606"  
/clone="IMAGE:347598"  
/clone\_lib="Soares\_fetal\_heart\_NbHH19W"  
/sex="unknown"  
/dev\_stage="19 weeks"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: heart; Vector: p7T3D (Pharmacia) with a  
modified polylinker; Site: 1: Not I; Site: 2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5'  
TGTTACCACTCTGAAGTGGAGCGGCCATCTTTTCTTTTCTTTT 3',  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified p7T3 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by  
M.Fatima Bonaldo. This library was constructed from the  
same fetus as the fetal lung library, Soares fetal lung  
NbHH19W."  
BASE COUNT 92 a 93 c 82 g 74 t 3 others  
ORIGIN  
Query Match 9.3%; Score 210; DB 10; Length 344;  
Best Local Similarity 99.4%; Pred. No. 1.le-75;  
Matches 310; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1551 ggtcttcgctaccccgagtagattttgataacattgaccccatatacaccaggca 1610  
Db 1 GGCTCTCGCTCACCCGAGTAGATATTTGCATACATTGTGACCCCATATACACCAAGGCA 60  
Qy 1611 aaaaatggctcctg9gcagcgagctatggatctggaacactgcaatccagtcagccatc 1670  
Db 61 AAAAAATGGCTCTGGGACGAGCGGTATGGGATCTGGAACACTGCAATCCAGTCAGCCATC 120  
Qy 1671 agatgcagactgaaattgcttacaggaatactctggtacagcagctgggtcccccct 1730  
Db 121 AGAGTGCAGACTGGAAATTTGTTACAGGAATTCCTGGCTACAGGAGCTTGGGTCCCCCT 180  
Qy 1731 cagttcttcagaaactgggacccgagctgggcacatgaacgactcacctcgtcaact 1790  
Db 181 CAGTCTTTACGACCTGGGACCCGACACNGTGGCAATGAACGGATCATCCTTGTCAACT 240

Qy 1791 ggcaaaagtgtatgctgttttcaacatcacagccgacccatagagaggtggacctatct 1850  
Db 241 GC AAAAGTGATGGCTTTTCAACATCACAGCGCCCATATGAGAGGTGGACCTATCT 300  
Qy 1851 aacaggtatcca 1862  
Db 301 AACAGGTATCCA 312  
RESULT 6  
BG661302 547 bp mRNA linear EST 05-JUL-2001  
LOCUS ia59e07.y1 Melton Human Islets H1Z1 Homo sapiens cDNA 5', mRNA  
DEFINITION sequence.  
ACCESSION BG661302  
VERSION BG661302.1 GI:13816937  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 547)  
REFERENCE Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,  
AUTHORS Lemishka, I., Scarce, M., Brestelli, J., Gradwohl, G., Clifton, S.,  
Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A.,  
Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas  
M., Gibbons M., McCann, R., Cole, R., Tsagaratishvili, R., Williams, T.,  
Jackson, Y. and Bowers, Y.  
TITLE Endocrine Pancreas Consortium  
JOURNAL Unpublished (2000)  
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
Endocrine Pancreas Consortium  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
MA 02138  
Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmelton@biohp.harvard.edu  
Library was constructed by Dr. Douglas Melton DNA sequencing by:  
Washington University Genome Sequencing Center For information on  
obtaining a clone please contact: Juliana Brown  
(brown@fas.harvard.edu)  
Seq primer: -40RP from Gibco  
High quality sequence stop: 482.  
FEATURES Location/Qualifiers  
1..547  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Melton Human Islets H1Z1"  
/sex="Both"  
/tissue.type="Islets of Langerhans"  
/dev\_stage="Adult"  
/lab\_host="TOP10"  
/note="Organ: Pancreas; Vector: pZErO-2; Site: 1: Not I;  
Site: 2: Xho I; Library constructed using SuperScript  
Plasmid Library kit (Life Technologies). cDNA made by  
oligo-dT priming. Xho I site destroyed during cloning.  
Size-selected by column fractionation; average insert  
size 1.59 kb. Primary library, unamplified."  
BASE COUNT 169 a 124 c 120 g 134 t  
ORIGIN  
Query Match 8.9%; Score 201; DB 10; Length 547;  
Best Local Similarity 100.0%; Pred. No. 4.5e-72;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2052 aaaaaaagaagaagaacagcagagaagcagctctcaggttcaacttgccattcaggttca 2111  
Db 157 AAAAAAGAGAGAAACAGCAGAGAAAGCACTCTCAGGTTCAACTTGCATTTCAGGTGTTA 216  
Qy 2112 cttgtggataagcacaataatttcctgttgggttaaacatttaacagttcttattcttca 2171

LOCUS BF218442 552 bp mRNA linear EST 06-NOV-2000  
DEFINITION 601881811F1 NTH\_MGC\_57 Homo sapiens cDNA clone IMAGE:4094211 5', mRNA sequence.





```

/db_xref="taxon:9606"
/clone="IMAGE:347598"
/clone_lib="Soares_fetal_heart_NBHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: heart; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dt) primer [5'
TGTTACCAATCTCAAGTGGGAGCGCGCATCTTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NBHH19W."

BASE COUNT      74 a   105 c   105 g   130 t
ORIGIN

Query Match      4.3%; Score 97; DB 10; Length 414;
Best Local Similarity 100.0%; Pred. No. 3.1e-29;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1625 ggcagcggctatggatctggaaactgcacatccagtcagcgcacatgcagcactg 1684
|||||
Db 413 GGCAGCGGCTATGGGATCTGGACACTGCAATCCAGTCAGCCATCAGATGCGAGCTG 354
|||||

QY 1685 gaaattgctacaggaatactcgtgctacagcactgg 1721
|||||
Db 353 GAAATTGCTACAGGAATCTCGGTCTACAGCGACTGG 317
|||||

RESULT 12
R43179/c
LOCUS      R43179      458 bp      mRNA      linear      EST 22-MAY-1995
DEFINITION y918d03.s1 Soares infant brain INIB Homo sapiens cDNA clone
IMAGE:32854 3' similar to SP:ARSB_HUMAN P15848 ARYLSULFATASE B
PRECURSOR ;, mRNA sequence.
ACCESSION  R43179.1 GI:825405
VERSION     EST.
KEYWORDS    human.
SOURCE      Homo sapiens
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 458)
AUTHORS     Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
,R., Williamson,A., Wohlmann,P. and Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
COMMENT     On May 8, 1995 this sequence version replaced gi:801403.
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1275
High quality sequence stops: 384 Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 1275 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 384.
FEATURES
source
1..458
/organism="Homo sapiens"
/db_xref="GDB:405201"

```

```

/db_xref="taxon:9606"
/clone="IMAGE:32854"
/clone_lib="Soares infant brain INIB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: whole brain; Vector: Laimid BA; Site_1: Not
I; Site_2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dt) primer [5'
AAGTGAAGAATTGCGCGCGCGCAATTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Laimid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT      74 a   110 c   119 g   152 t
ORIGIN

Query Match      4.1%; Score 94; DB 10; Length 458;
Best Local Similarity 100.0%; Pred. No. 5.2e-28;
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1881 ctacggggctctcacagttcaacaaactgcagtcgcggtcaggtatccccccaaagac 1940
|||||
Db 179 CTACGGAGGCTCTCACAGTTCAACAAACTGCAGTGGCGGTATCCCCCAAGAC 120
|||||

QY 1941 ccagaagtaaccctagctcaatggagggtct 1974
|||||
Db 119 CCCAGAAGTACCCTAGGCTCAATGGAGGGGTCT 86
|||||

RESULT 13
AA159953/c
LOCUS      AA159953      463 bp      mRNA      linear      EST 16-DEC-1996
DEFINITION z075a09.s1 Stratagene pancreas (#937208) Homo sapiens cDNA clone
IMAGE:592696 3' similar to SW:ARSB_HUMAN P15848 ARYLSULFATASE B
PRECURSOR ;, mRNA sequence.
ACCESSION  AA159953
VERSION     AA159953.1 GI:1734444
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 463)
AUTHORS     Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins
,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaskis,E.,
Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 369.
FEATURES
source
1..463
/organism="Homo sapiens"
/db_xref="GDB:4624082"
/clone="IMAGE:592696"
/lab_host="Stratagene pancreas (#937208)"

```

/note="Organ: pancreas; Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Pancreatic adenocarcinoma cell line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'."

BASE COUNT 82 a 115 c 115 g 151 t

ORIGIN

Query Match 4.18; Score 94; DB 9; Length 463;  
Best Local Similarity 100.0%; Pred. No. 5.1e-28;  
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1881 ctacgaggtctcacagtccaacaaactcagtcagtcggtcaggtatcccccaagaac 1940

Db 172 CTACGGAGGCTCTCACAGTTCAACAAACTCGAGTCCGGTCAGGTATCCCCCAAGAC 113

Qy 1941 cccgaagaatcaccttagtctcaatggaggggtct 1974

Db 112 CCCAGAAGTAACCTAGGCTCAATGGAGGGTCT 79

RESULT 14  
BF696654 602124759f1 NIH\_MGC\_56 Homo sapiens cDNA clone IMAGE:4281567 5',  
LOCUS mRNA sequence.  
DEFINITION

ACCESSION BF696654

VERSION BF696654.1 GI:11982062

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 849)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: CLONETECH Laboratories, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LICM110 row: d column: 16

High quality sequence stop: 579.

Location/Qualifiers

FEATURES

source

1..849

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4281567"

/clone\_lib="NIH\_MGC\_56"

/tissue\_type="primitive neuroectoderm"

/lab\_host="DH10B (T1 phage-resistant)"

/note="Organ: brain; Vector: PDNR-LIB (Clontech); Site\_1:

Sf11 (99ccctcgccgc); Site\_2: Sf11 (99ccctcgccgc);

Double-stranded cDNA was prepared from cell line RNA.

5' and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor

sequence: 5'-ATTCTAGAGCGCGCGGCACATG-dt(30)BN-3'

(where B = A, C, or G and N = A, C, G, or T). Average

insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies

contained inserts by PCR. This library was enriched for

full-length clones and was constructed by Clontech

Laboratories (Palo Alto, CA)."

BASE COUNT 237 a 175 c 180 g 257 t

ORIGIN

Query Match 3.6%; Score 81; DB 10; Length 849;  
Best Local Similarity 100.0%; Pred. No. 9.1e-23;  
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2060 aagaagaacacagcagaagcagtcctcaggttcaacttcgacctcaggtgttgaga 2119

Db 1 AAGAAGAAACAGCAGAAGCAGTCTCAGGTTCAACTTCCCATTCAGGTGTGTGGA 60

Qy 2120 taagcacaaatatttcctgtt 2140

Db 61 TAAGCACAAATATTCCTGTT 81

RESULT 15

AA159952

LOCUS

DEFINITION

2075a09.r1 Stratagene pancreas (#937208) Homo sapiens cDNA clone

IMAGE:592696 5', mRNA sequence.

AA159952

ACCESSION

VERSION

KEYWORDS

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,

Chissee, S., Dietrich, N., DuBuque, T., Lacy, M., Le, N., Mardis, E., Moore

, M., Hultman, M., Kucaba, T., Prange, C., Rifkin, L., Rohlfing, T.,

Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E.,

Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags

Genome Res. 6 (9), 807-828 (1996)

97044478

TITLE

JOURNAL

MEDLINE

COMMENT

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 274.

FEATURES

source

1..407

/organism="Homo sapiens"

/db\_xref="GDB:4624082"

/db\_xref="taxon:9606"

/clone="IMAGE:592696"

/clone\_lib="Stratagene pancreas (#937208)"

/lab\_host="SOLR cells (kanamycin resistant)"

/note="Organ: pancreas; Vector: pBluescript SK-; Site\_1:

EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer:

Oligo dT. Pancreatic adenocarcinoma cell line. Average

insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor

sequence: 5' GAATTCGGCAGGAG 3' -3' adaptor sequence: 5'

CTCGAGTTTTTTTTTTTTTTT 3'."

BASE COUNT 82 a 99 c 114 g 104 t

ORIGIN

Query Match 3.1%; Score 70; DB 9; Length 407;

Best Local Similarity 100.0%; Pred. No. 4.1e-18;

Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 83 gagagaggagaaagtgaatgtctgtgagaagcagcgcctcttcttccggag 142

Db 215 GAGGAGGAGGAGAAAGTGAATGTCTGGAGAAGCGAGCCCTCTCTTCTCCGGAG 274

Qy 143 tccatccat 152

ORIGIN

Db 275 TCCCATCCAT 284

Search completed: July 23, 2002, 17:46:44  
Job time: 9520 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.  
OM protein - protein search, using sw model  
Run on: July 23, 2002, 17:46:49 : Search time 133.2 Seconds  
(without alignments)  
714.318 Million cell updates/sec

Title: US-09-495-823-7  
Perfect score: 3012  
Sequence: 1 MAPRCAGHPPPPSPQACVC.....VPVRYPPKDRSPNRLNGV 550  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Database : SPTREMBL\_19:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1293.5	42.9	503	5 Q9NJU8	Q9nj8 helix pomat
2	920	30.5	542	5 Q9V6F2	Q9v6f2 drosophila
3	846.5	28.1	579	5 Q9VVM4	Q9vvm4 drosophila
4	752	25.0	996	5 Q9VVK9	Q9vvk9 drosophila
5	646.5	21.5	266	5 Q9NJU7	Q9nju7 helix pomat
6	556	18.5	486	5 Q9VVM1	Q9vvm1 drosophila
7	507	16.8	169	4 Q9H5D8	Q9h5d8 homo sapien
8	472	15.7	521	16 Q9A922	Q9a922 caulobacter
9	447	14.8	520	11 Q9JHK9	Q9jkh9 mus musculus
10	443	14.7	520	11 Q9QK08	Q9qk08 mus musculus
11	432	14.3	554	3 Q9C0V7	Q9c0v7 schizosacch
12	395.5	13.1	537	16 Q92UC0	Q92uc0 rhizobium m
13	393	13.0	555	16 Q92YF6	Q92yf6 rhizobium m
14	393	13.0	577	2 Q9X759	Q9x759 klebsiella
15	391.5	13.0	525	11 Q9D3B4	Q9d3b4 mus musculus
16	391	13.0	577	2 Q9L4Y4	Q9l4y4 klebsiella

17	391	13.0	577	2 Q9L4Y2	Q9l4y2 klebsiella
18	384.5	12.8	525	4 Q96EG1	Q96eg1 homo sapien
19	383	12.7	525	4 Q9Y2K4	Q9y2k4 homo sapien
20	378	12.5	509	4 Q96CJ0	Q96cj0 homo sapien
21	377	12.5	557	16 Q98BQ3	Q98bq3 rhizobium l
22	371	12.3	506	11 Q9DC66	Q9dc66 mus musculus
23	360	12.0	475	2 Q9F821	Q9f821 escherichia
24	359	11.9	559	5 Q16138	Q16138 helicobacter
25	355.5	11.8	551	5 Q25384	Q25384 hemocentrot
26	344.5	11.4	303	5 Q18337	Q18337 caenorhabdi
27	336	11.2	545	16 Q92ZH5	Q92zh5 rhizobium m
28	321.5	10.7	517	2 Q9L5W0	Q9l5w0 prevotella
29	319.5	10.6	787	16 P95059	P95059 mycobacteri
30	285	9.5	512	2 Q93P98	Q93p98 microscilla
31	269	8.9	453	16 Q9CKE0	Q9cke0 pasteurella
32	256	8.5	452	5 Q18924	Q18924 caenorhabdi
33	251	8.3	649	10 Q9ATG5	Q9atg5 chlamydomon
34	248	8.2	382	4 Q9UHJ8	Q9uhj8 homo sapien
35	243.5	8.1	970	16 Q65931	Q65931 mycobacteri
36	236	7.8	524	5 Q9VE24	Q9ve24 drosophila
37	230	7.6	249	2 Q93P96	Q93p96 microscilla
38	221.5	7.4	503	16 Q9I799	Q9i799 pseudomonas
39	219.5	7.3	709	5 Q21376	Q21376 caenorhabdi
40	218	7.2	430	2 Q9KYR2	Q9kyr2 drosophila
41	218	7.2	430	2 Q9KYR2	Q9kyr2 streptomyce
42	213.5	7.1	787	16 Q06776	Q06776 mycobacteri
43	212	7.0	538	16 Q9I1E5	Q9i1e5 pseudomonas
44	207	6.9	202	2 Q93P97	Q93p97 microscilla
45	206.5	6.9	867	13 Q90XB6	Q90xb6 coturnix co

ALIGNMENTS

RESULT 1

Q9NJU8	ID	Q9NJU8	PRELIMINARY;	PRT;	503 AA.
AC	Q9NJU8;				
DT	01-OCT-2000	(Tremblrel. 15, Created)			
DT	01-OCT-2000	(Tremblrel. 15, Last sequence update)			
DT	01-DEC-2001	(Tremblrel. 19, Last annotation update)			
DE	SULFATASE 1.				
GN	SULF1.				
OS	Helix pomatia (Roman snail) (Edible snail).				
OC	Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;				
OC	Helicacea; Helicidae; Helix.				
OX	NCBI_TaxID=6536;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=20233298; PubMed=10772344;				
RA	Wittstock U., Fischer M., Svendsen I., Halkier B.A.;				
RT	"Cloning and characterization of two cDNAs encoding sulfatases in the				
RL	Roman snail, Helix pomatia."				
RL	IUBMB Life 49:71-76(2000).				
DR	EMBL; AF109924; AAF30402.1; -.				
DR	HSSP; P15848; IFSU.				
DR	InterPro: IPR000917; Sulfatase.				
DR	Pfam: PF00884; Sulfatase; 1.				
DR	PROSITE; PS00149; SULFATASE_2; 1.				
SQ	SEQUENCE 503 AA; 55821 MW; AC45334BA694413D CRC64;				

Query Match 42.9% Score 1293.5; DB 5; Length 503;  
Best Local Similarity 48.3%; Pred. No. 8.7e-102;  
Matches 25; Conservative 82; Mismatches 141; Indels 51; Gaps 8;

QY	37	CLLFTYGLSWGQALLEEBCGALLAQAGKLEPSTTSQPHLIFILADDQGRDVGVHGS	96
Db	4	CLLVLIATACAVADQSS---ASACTROD-----AGOPNIVFLADDFGHDVGVHGS	54
QY	97	EIKPTTLDKLAAGVKNLENYVQICTPSPRSQFTIGYQIHTGLQHSIRTPQNCPLD	156
Db	55	EIHTPTLDALSASGVRLNYYVQICTPSTRSOLMSGRYQIHTGLQHSIRTPQNCPLD	114



DT	01-MAY-2000	(TReMBLrel. 13, Last sequence update)
DE	01-JUN-2001	(TReMBLrel. 17, Last annotation update)
DT	CG7402	PROTEIN.
GN	CG7402	
OS	Drosophila melanogaster	(Fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
OC	Pterygota; Neoptera; Endopterygota; Diptera; brachycera; Muscomorpha;	
OC	Ephydroidea; Drosophilidae; Drosophila.	
ON	NCBI_Taxid=7227;	
OX	[1]	
RC	SEQUENCE FROM N.A.	
RP	STRATN-BERKELEY.	
RX	MEDLINE=20196006; PubMed=10731132;	
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,	
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	
RA	Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,	
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,	
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,	
RA	Abriil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,	
RA	Ballwey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,	
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,	
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,	
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,	
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,	
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,	
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,	
RA	Durbin K.J., Evangelista A.C., Ferraz C., Ferriera S., Fleischmann W.,	
RA	Foster C., Gabrielian A.E., Garcia N.S., Gelbart W.M., Glasser K.,	
RA	Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,	
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,	
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,	
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,	
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,	
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,	
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,	
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,	
RA	Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,	
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,	
RA	Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,	
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,	
RA	Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,	
RA	Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,	
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,	
RA	Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissensbach J.,	
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,	
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,	
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,	
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;	
RT	The genome sequence of Drosophila melanogaster."	
RL	Science 287:2185-2195(2000).	
DR	EMBL; AE003522; AAF49287.1; -.	
DR	HSSP; P15848; 1FSU.	
DR	FlyBase; FBgn0036768; CG7402.	
DR	InterPro; IPR000917; Sulfatase.	
DR	Pfam; PF00884; Sulfatase; 1.	
DR	PROSITE; PS00149; SULFATASE 2; 1.	
SQ	SEQUENCE 579 AA; 65053 NW; EBF2F4F13D759298 CRC64;	

	Query Match	28.1%	Score	846.5;	DB 5;	Length	579;			
	Best Local Similarity	36.5%;	Pred.	No. 1.5e-63;						
	Matches	192;	Conservative	85;	Mismatches	192;	Indels	57;	Gaps	14;
QY	73	TSQPHLFIILADGGRDVGHGHS-EKTPPLDKLAEGVKLENYYVQIPTCPSPRSQFIT	131							
Dd	:	:::::	:	:	:	:	:	:	:	:
	25	STRKNIVIIITDDMGMDNVDFHGSNQILTPNIDALAYNGILLNRHYVNPCLCTPSPRATLLT	84							
QY	132	GKYOIHTGLQHSIIRTPQNCLPLDNATLPKLKEVGYSTHWCKWHLGFYRKCEMPTFR	191							
Dd	:	:::::	:	:	:	:	:	:	:	:
	85	GKPIHTGMGHFVIITDEPWGLPORERLMPEIFNDAGYSTHLVGKWHLGFWRKKDTPTMR	144							
QY	192	GFDTFFGSLLSGGDYTHY--KCDSPGMCGDYDLVENDNAAMDYDNGIYSTOMYTORVOOI	249							

Db	145	GFPHFGFYNGYIDYDQVRLMDRNYSAGLD-FRDLEPCPEANGTYATEFTSEAKRI	203
Qy	250	LASHNPTKFLFYIAYOAVH-----SPLQAPGRFYFHYRSIININRRRYAAMLSCLEDAI	304
Db	204	IEQHDKSPLFLFWLVLAVITGNEDSFMQAEVAKFPHIRDPKRTYAGMISLSDKSV	263
Qy	305	NNVTALKTYGFYNNSSIIYSSDNGQPTAG-----GSMNPLRGSKGTWEGGIRAVGV	359
Db	264	AQTIGALKDNGMLNLSIILYSDNGA-PTGIHNSAGSNYPYRGQKESPWEGGIRSAGAL	322
Qy	360	HSPLLKKNKTYCKEPVHITDNYPTLISLABQIDEDIQDGDYDIWETISEGLRSPV-DI	418
Db	323	WSPLLKRGVYSNQAIHAVDWLPTLAGAGVSUPQDPLDGLINLWPLMS-GNPEPKPRTM	381
Qy	419	LHNIDPIY-----TKAKNGSWAAG-YGIW-----NTAIQSAIRVQHVKL	456
Db	382	IHVLEDEVGYSSYMRDTFLKTVNGSSSPKGRYDQWLGELETNEDDPLGESYEQHVLSADVOS	441
Qy	457	LTGNPGY-----SDWVPPQSFSLGNPRNWHNERITSSTGSKSVMLFNITADPYE	504
Db	442	LLGNRLTKDRIQRMRSEATETCPPIEGONPLESHFKCEPL-----KAPCFDLAKDPCE	496
Qy	505	RVDSLNRYPGIVKLLRRLRSQFNKTAVP-VRYPPKDPDRSNPRLNG	549
Db	497	RYNLAQMYPLQLOQLQADELQIRKTAIPSAVRPHSDSRANPTFHNG	542
RESULT	4		
Q9VVVK9	AD	Q9VVK9 PRELIMINARY; PRT; 996 AA.	
IC	Q9VVK9		
DT	01-MAY-2000	(T-EMBLrel. 13, Created)	
DT	01-MAY-2000	(T-EMBLrel. 13, Last sequence update)	
DT	01-JUN-2001	(T-EMBLrel. 17, Last annotation update)	
DE	CG5584	PROTEIN.	
GN	CG5584		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Prervgota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxId=7227;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BERKELEY.		
RC	MEDLINE=20196006; PubMed=107311132;		
RA	Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Ananietis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,		
RA	George R.A., Lewis S.E., Richards S., Ashburner S., Henderson S.N.,		
RA	Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,		
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,		
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,		
RA	Baril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,		
RA	Bailey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,		
RA	Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,		
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,		
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,		
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,		
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,		
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,		
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,		
RA	Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,		
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,		
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,		
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,		
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,		
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,		
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,		
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,		
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,		
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,		
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,		
RA	Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,		





Db 394 LFDIEADPCERSNLVAEYQNSTIFLDLSRIQQFAKQAHPPNNKPGDPNCDPR 446

```

RESULT 7
ID Q9HSD8 PRELIMINARY; PRT; 169 AA.
AC Q9HSD8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CDNA: FLJ23548 FIS, CLONE LNC08487.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,
RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,
RA Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Ota T., Suzuki Y.,
RA Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,
RA Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK07201; BAB15689.1; -.
DR HSSP; P15848; 1FSU.
SQ SEQUENCE 169 AA; 18905 MW; 4583C92FFCC3380 CRC64;

Query Match 16.8%; Score 507; DB 4; Length 169;
Best Local Similarity 96.9%; Pred. No. 2.2e-35;
Matches 93; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 455 KLLTGNPGYSQWVPQSFNSLGNPRNWHNERTISTGKSWLFINITADPYRVDLSNRYPG 514
Db 25 RLLTGNPGYSQWVPQSFNSLGNPRNWRNERTITLSTGKSWLFINITADPYRVDLSNRYPG 84
Qy 515 IVKLLRLSOFNKTAVPVRYPPKDPSPNRLNGGV 550
Db 85 IVKLLRLSOFNKTAVPVRYPPKDPSPNRLNGGV 120

RESULT 8
ID Q9A922 PRELIMINARY; PRT; 521 AA.
AC Q9A922;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SULFATASE FAMILY PROTEIN.
GN CC1172.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OC NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Niernan W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K., O.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Shapero L., Fraser C.M.;
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AA005795; AAK23156.1; -.
DR HSSP; P15848; 1FSU.
DR TIGR; CC1172; -.
DR InterPro; IPR000917; Sulfatase.
DR Pfam; PF00884; Sulfatase; 1.
DR PROSITE; PS00523; SULFATASE 1; 1.

```







Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H., Kuehl P., Lewis S., Matsuo Y., Nikolaï D.I., Pesole G., Quackenbush J., Schriml L.M., Staubili F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hall D., Hofman M., Humé D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L.L., Mashima J., Mazzarelli J., Mommaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasakí H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S., Hayashizaki Y.	
R	*Functional annotation of a full-length mouse cDNA collection.*;
RT	Nature 409:685-690(2001).
DR EMBL:	AK018132; BAB31086.1; -
DR HSSP:	P15289; LAUK
DR MGD:	MGI:1921258; 633040G.P08RIK.
DR InterPro:	IPIR0000917; Sulfatase.
DR Pfam:	PF00884; Sulfatase; 1.
DR PROSITE:	PS00149; SULFATASE_2; 1.
SQ SEQUENCE	525 AA; 57449 MW; DF889E39866DIEIF CRC64;

  

Query Match      13.0%;    Score 391.5;   DB 11;   Length 525;  
Best Local Similarity   26.7%;   Pred. No. 8.7e-25;  
Matches 138; Conservative   89; Mismatches 168; Indels 121; Gaps

---

Y QY	71	TSTSQPHLIFFILLADDOGFDDVGYHGHSEIK-TPTLDKLAAEGVKLENYY-VQPICTPSRSQ	128
D b	31	TRAPQNIVIIILADDGWGLGANNAETKDFTNLDKMASRGMRFDVFHAASTCSPSRA	90
Y QY	129	FITGKYIHTQLSHSIIRTPQNCPLPONATIPORLKEGVYSTHMVGKWHLGGFYRKECMP	188
D b	91	LTTGRGLRNGVTNH-FAVTSYGVLPNETTLLAEVLRQGYYTAMIGKWHLG-HHGSYHP	148
Y QY	189	TRRGFTTFPG---SLLGSDGY--YTHYKCD-----SPGMCGY----DLYEEND	227
D b	149	NFRGFDFYFGIPYSNDMGCTDPAGYNYPCCACPORQDGLWRNPGRDCYTIDVALPLYENLN	208
Y QY	228	AAWDYDNGIYSTQMRYTOR-VOOILASHNPTKFIIFYIAVAHVSLQ-APGRYFEHYRSI	285
D b	209	IVEQPVNLISGLAQKAYEAERVEIFEQAOSTSGRFELLYGOAHMHVPLSVTPPLAHPQROSL	268
Y QY	286	INTNRRRYAAMLSCDELAINNNYLTLKTYGFYNNNSIIIYSSDNGOPTA-----GGSNWP	340
D b	269	YRASLEMDSLVCQIKDKVDHYA-----RENTLLWFTGDNG-PWAQKCELAGSVGP	318
Y QY	341	LRG-----SKGTYWEEGIRAVGFVHSPLLKNKGTVCKPEVHIIT-----DWPP	382
D b	319	FFGLWOTHOGGSPTKTOTTWEGGHRPALAYWP-----GRV---PANVTSTALLSLDIFF	370
Y QY	383	TLTSLAEGOIDEIOLDGWDIWTETISEGLRSRPVDILHNIDIPTYKANG-----SW	434
D b	371	TVTALAGASLPNPRKFDDRVDSVV-----LFQKSOMGRHFVFPHPMSG	412
Y QY	435	AAG-YGIWNITAQS AIRVHWK-LLTGNPGYSDVWPPQSFNSLGNPKRWNIERTISTS	491
D b	413	AGEYGV---ALQT-VRLNHYAFYITGAACAD-----GSVGEQHVVAPL-----	454
Y QY	492	SVWLFNITADPYERVDL-----SNRYPGVIKKLLRRLS	524
D b	455	--INFNDADEAGMFLQKGSPEYQEVLQQVYRALA	487

Search completed: July 23, 2002, 19:06:40  
Job time: 4791 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 17:07:14 ; Search time 79.69 Seconds  
(without alignments)  
766.603 Million cell updates/sec

Title: US-09-495-823-7  
Perfect score: 3012  
Sequence: 1 MAPRCAGHPPPPSPQACVC.....VPRYPKDPKPSNRLNGGV 550

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2996	99.5	599	22 AAB85483	Human 23553 sulfat
2	2307.5	76.6	515	20 AAV41701	Human PRO708 prote
3	2307.5	76.6	515	21 AAB44257	Human PRO708 (ONQ3
4	2307.5	76.6	515	22 AAU29061	Human PRO polypept
5	1859.5	61.7	569	22 AAB85482	Human 25278 sulfat
6	1562	51.9	289	22 AAU14421	Human novel protei
7	1415.5	47.0	533	22 AAB51184	Human sulfatase pr
8	1079	35.8	366	22 AAU14185	Human novel protei
9	989	32.8	187	22 AAU25415	Human protein sequ
10	920	30.5	542	22 ABB63962	Drosophila melanog
11	846.5	28.1	579	22 ABB63246	Drosophila melanog

12	752	25.0	996	22 ABB61654	Drosophila melanog
13	603.5	20.0	128	22 AAU23335	Novel human enzyme
14	556	18.5	486	22 AAB63259	Drosophila melanog
15	457.5	15.2	492	22 ABB52501	Escherichia coli p
16	443	14.7	520	22 AAB85485	Pfam consensus seq
17	429	14.2	552	22 AAB85484	Pfam consensus seq
18	424.5	14.1	590	22 AAU78688	Human protein SEQ
19	422.5	14.0	522	22 AAU41939	Human polypeptide
20	413	13.7	583	22 AAX39920	Human steroid sulp
21	404	13.4	583	22 AAB51185	Human sulfatase pr
22	402	13.3	646	22 AAU41129	Human polypeptide
23	399	13.2	631	22 ABG22561	Novel human diagno
24	394.5	13.1	589	22 AAU39343	Human polypeptide
25	384.5	12.8	525	22 AAB85480	Human 22438 sulfat
26	384.5	12.8	552	22 ABB12445	Human bone marrow
27	378.5	12.6	525	22 AAU12199	Human PRO1777 poly
28	356	11.8	507	22 AAB51183	Human sulfatase pr
29	353.5	11.7	736	22 ABG22560	Novel human diagno
30	336.5	11.2	551	22 AAB51186	Sea urchin sulfata
31	324	10.8	571	22 AAG98936	E. coli growth and
32	319.5	10.6	787	22 AAG81127	Mycobacterium tube
33	255	8.5	1120	22 ABG07425	Novel human diagno
34	248.5	8.3	552	22 AAU39414	Human polypeptide
35	243.5	8.1	602	22 AAU41200	Human polypeptide
36	236	7.8	524	22 ABB69201	Drosophila melanog
37	236	7.8	625	22 ABG20519	Novel human diagno
38	221	7.3	510	22 AAB51182	Human sulfatase pr
39	219.5	7.3	709	22 AAE00435	C. elegans sulfata
40	216	7.2	1689	22 ABG16013	Novel human diagno
41	216	7.2	2417	22 ABG28727	Novel human diagno
42	214.5	7.1	822	22 ABG28571	Novel human diagno
43	214	7.1	502	20 AAU42790	Human sulphamidase
44	214	7.1	502	20 AAU90233	Human sulphamidase
45	214	7.1	502	22 AAE08548	Human sulphamidase

ALIGNMENTS

RESULT	1
AAB85483	
ID	AAB85483 standard; Protein; 599 AA.
AC	AAB85483;
XX	
DT	25-SEP-2001 (first entry)
XX	
DE	Human 23553 sulfatase polypeptide.
XX	
KW	Sulfatase; 23553 sulfatase; nootropic; neuroprotective; antibacterial;
KW	antiinflammatory; vasotrophic; antitumor; gene therapy; human.
XX	
OS	Homo sapiens.
XX	
PN	WO200155411-A2.
XX	
PD	02-AUG-2001.
XX	
PF	31-JAN-2001; 2001WO-US03266*
XX	
PR	31-JAN-2000; 2000US-0495823.
XX	
PA	(MILL-) MILLENNIUM PHARM INC.
XX	
PI	Glucksmann MA, Williamson M, Rudolph-Owen LA, Tsai F;
XX	
DR	WPI; 2001-476214/51.
DR	N-PSDB; AAH46863, AAH46867.
XX	
PT	Novel human sulfatase polypeptides useful for treating and diagnosing
PT	sulfatase-related disorders such as cerebrovascular diseases, acute
PT	meningitis, multiple sclerosis, degenerative diseases and tumor
XX	

PS Claim 9; Fig 15; 180pp; English.

XX The invention provides 22438, 23553, 25278 or 26212 human sulfatase polypeptides and polynucleotides. The sulfatase genes and polypeptides are useful for treating disorders involving the brain such as cerebrovascular diseases, infections such as acute meningitis, demyelinating diseases including multiple sclerosis, degenerative diseases affecting the cerebral cortex including Alzheimer's disease and Pick disease, spinocerebellar degenerations including spinocerebellar ataxias including Friedreich ataxia, and ataxia telangiectasia, CC degenerative diseases affecting motor neurons including amyotrophic lateral sclerosis, inborn errors of metabolism such as leukodystrophies, CC toxic and acquired metabolic diseases, including vitamin deficiencies, CC and neurocutaneous syndromes (phakmatoses) including neurofibromatosis. CC The present sequence represents a human 23553 sulfatase polypeptide.

XX Sequence 599 AA;

Query Match 99.5%; Score 2996; DB 22; Length 599;  
Best Local Similarity 99.6%; Pred. No. 5.3e-276;  
Matches 548; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MAPRGCAHPPPPSPQACVCPCKMLAMGALAGFWILCLLTGYLSWGOALEEEEGALLA 60  
Db 1 maprgcaghppppspqacvcpckmlamgalagfwilclltygylswgaleeeegalla 60

Qy 61 QAGEKLEPSTSTSQPHLIFLILADDOGFRDYGHGSEIKTPTLKLAAEGVKLENYVQP 120  
Dy 61 qageklepststsqphlfliladdqgfrdyghgseiktpklklaaegvklennyvqp 120

Qy 121 ICTPSRSQIFGKQIHTGLQHSIRTPQNCPLDNATLPQKLKEVGYSHVMVGKWHLG 180  
Dy 121 ictpsrqifgkqihgtglqhsirtpqncpldnatlpqklkevgyshvmvgkwhlg 180

Qy 181 FYRKECMTPTRRGFTDFGSLGSGDYTHYKCDSPGCGYDLYENDNAWDYDNGIYSTQ 240  
Dy 181 fyrkecmptrrgftdfgslgsgdythykcdspgcmgydlyendnaawdydngiystq 240

Qy 241 MYTORVQOILASHNPTKPIFLYIAQVHSPLOAPGRYFEHYRSTIINRRYAAMLSCL 300  
Dy 241 mytorvqoqlashnptkpfilyiaqvahsploapgrfyehyrstlinrryaaamlsc 300

Qy 301 DEANNVTALKTYGFYNNSIIYSSDNGQPTAGGSNPLRGSKGTWEGGIRAVGFVH 360  
Dy 301 deannvtlalktygfynnsiiyssdngqptaggsnplrgskgtweggiravgfvh 360

Qy 361 SPLLNKGTVCKEPVHTDWTPTLISAEQIDEDIQDGDYDIWETISEGLRSPRVDLH 420  
Dy 361 spllnkgtvckevhtdwtptlisaegqidediqdgydiwetiseglrsprvdlh 420

Qy 421 NIDPTYTKAKGNSAAGYGIWNTAQSAIRVQHWKLLTGNPGYSDWPPQSFSLGNPNRW 480  
Dy 421 nidplytkakngswaagygiwntaqsaairvqhwwklltgnpgysdwppqsfslgnp 480

Qy 481 HNERITSTGKSVMLFNITADPYERVDLSNRPYGVIRKLLRRLSQFNKTAVPVPPKDP 540  
Dy 481 hneritstgksvmlfnitadpyervdlsnrypgvirklrrlsqfnktavpvpkdp 540

Qy 541 RSNPRLNGGV 550  
Dy 541 rsnprlnggv 550

RESULT 2  
AA41701  
ID AAY41701 standard; Protein; 515 AA.  
XX  
AC AAY41701;  
XX  
DT 07-DEC-1999 (first entry)  
XX Human PRO708 protein sequence.

XX Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;  
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;  
KW secreted protein; transmembrane protein.  
OS Homo sapiens.  
PN WO9946281-A2.  
DD 16-SEP-1999.  
XX 08-MAR-1999; 99WO-US05028.  
PR 10-MAR-1998; 98US-0077450.  
PR 11-MAR-1998; 98US-0077632.  
PR 11-MAR-1998; 98US-0077641.  
PR 11-MAR-1998; 98US-0077649.  
PR 12-MAR-1998; 98US-0077791.  
PR 13-MAR-1998; 98US-0078004.  
PR 17-MAR-1998; 98US-0040220.  
PR 20-MAR-1998; 98US-0078886.  
PR 20-MAR-1998; 98US-0078910.  
PR 20-MAR-1998; 98US-0078936.  
PR 20-MAR-1998; 98US-0078939.  
PR 25-MAR-1998; 98US-0079294.  
PR 26-MAR-1998; 98US-0079656.  
PR 27-MAR-1998; 98US-0079663.  
PR 27-MAR-1998; 98US-0079664.  
PR 27-MAR-1998; 98US-0079689.  
PR 27-MAR-1998; 98US-0079728.  
PR 27-MAR-1998; 98US-0079786.  
PR 30-MAR-1998; 98US-0079920.  
PR 30-MAR-1998; 98US-0079923.  
PR 31-MAR-1998; 98US-0080105.  
PR 31-MAR-1998; 98US-0080107.  
PR 31-MAR-1998; 98US-0080165.  
PR 31-MAR-1998; 98US-0080194.  
PR 01-APR-1998; 98US-0080327.  
PR 01-APR-1998; 98US-0080328.  
PR 01-APR-1998; 98US-0080333.  
PR 01-APR-1998; 98US-0080334.  
PR 08-APR-1998; 98US-0081049.  
PR 08-APR-1998; 98US-0081070.  
PR 08-APR-1998; 98US-0081071.  
PR 09-APR-1998; 98US-0081195.  
PR 09-APR-1998; 98US-0081203.  
PR 09-APR-1998; 98US-0081229.  
PR 15-APR-1998; 98US-0081817.  
PR 15-APR-1998; 98US-0081838.  
PR 15-APR-1998; 98US-0081952.  
PR 15-APR-1998; 98US-0081955.  
PR 21-APR-1998; 98US-0082568.  
PR 21-APR-1998; 98US-0082569.  
PR 22-APR-1998; 98US-0082700.  
PR 22-APR-1998; 98US-0082704.  
PR 22-APR-1998; 98US-0082804.  
PR 23-APR-1998; 98US-0082767.  
PR 23-APR-1998; 98US-0082796.  
PR 27-APR-1998; 98US-0083336.  
PR 28-APR-1998; 98US-0083342.  
PR 29-APR-1998; 98US-0083352.  
PR 29-APR-1998; 98US-0083495.  
PR 29-APR-1998; 98US-0083496.  
PR 29-APR-1998; 98US-0083499.  
PR 29-APR-1998; 98US-0083500.  
PR 29-APR-1998; 98US-0083545.  
PR 29-APR-1998; 98US-0083554.  
PR 29-APR-1998; 98US-0083558.  
PR 29-APR-1998; 98US-0083559.  
PR 30-APR-1998; 98US-0083742.  
PR 05-MAY-1998; 98US-0084366.  
PR 06-MAY-1998; 98US-0084414.  
PR 06-MAY-1998; 98US-0084441.





PT to target bioactive molecules to specific cells, and to modulate  
XX cellular activities -  
PS Claim 12; Fig 43; 636pp; English.  
XX  
CC AAC78458 to AAC78599 represent polynucleotide and EST (expressed  
CC sequence tag) sequences which encode secreted or transmembrane PRO  
CC polypeptides. The PRO polynucleotides and polypeptides have cytostatic  
CC activity. The polynucleotides and polypeptides can be used for detecting  
CC the presence of PRO polypeptides in samples, for linking bioactive  
CC molecules to cells and for modulating biological activities of cells,  
CC using the polypeptides for specific targeting. The polypeptide targeting  
CC can be used to kill the target cells, e.g. for the treatment of cancers.  
CC The polypeptide pairs provide specific targeting of bioactive molecules  
CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in  
CC the isolation of the PRO polynucleotide sequences.  
XX  
SQ Sequence 515 AA;  
  
Query Match 76.6%; Score 2307.5; DB 21; Length 515;  
Best Local Similarity 93.7%; Pred. No. 1.5e-210;  
Matches 434; Conservative 3; Mismatches 15; Indels 11; Gaps 4;  
  
Qy 1 MAPRCAGHPPPSPQACVCPKMLAMGALAGFWILCLLTGYLSWGQALFEEEGALLA 60  
Db 1 maprgcaghhppspqacvcpqkmlangalagfwilclltgyllswgqaleeeegalla 60  
  
Qy 61 QAGKLEPSTTSTQPHLIFLADQGFQVYHGSEIKTPTLKLAAEGVKLENYYVQP 120  
Db 61 qageklepsttstqphlilfildqgfrdvgyhgseiktptldklaaegvklenyyvqp 120  
  
Qy 121 ICTPSRSQFITGKQIHTGLQHSIRTPQNCPLDNLATLPQKLKEVCYSTHMYGKWHLG 180  
Db 121 ictpsrsqfitygkqihtglqhsirtpqncpldnlatlppqklkevgysthmygkwhlg 180  
  
Qy 181 FYRKECMPTRRGFTFFGSLGSGDYTHYKCDSPGCMGYDLYENDNAWDYDNGIYSTQ 240  
Db 181 fnrkecmptrrgftffgslgsgdythykcdspgcmgydlyendnaawdydngiystq 240  
  
Qy 241 MYTORVOOILASHNPTKPIFYLYAQAVHSPLOAPGRYFEHYRSIININRRRYAAMLSC 300  
Db 241 mytrvqqlashnptkpiilyayqavhsploapgrfyehyrsilininrrryaamlsc 300  
  
Qy 301 DEANNVTALKTYGFYNNSSIIYSSONGQPTAGGSNWPGRSGKTYWEGIRAVGFVH 360  
Db 301 deannvtalktygfynnsiiyssongqptaggsnwpgrsgktywegiravgfvh 360  
  
Qy 361 SPLLNKGTVCKEPVHITDWTYPTLISLAEGQIDEDIQLDGYDIWETISEGLRSPRVDILH 420  
Db 361 spllnkgtvckevhitdwtptyptlislsegqidediqldgydiwetiseglrsprvdlh 420  
  
Qy 421 NIDPIYT---KAKNGSWAAGYGIWNTAIQSAIRVQHKKLLTGN 460  
Db 421 nidp-tytrqkmapqgqamsgs---tlqss---qpsecstgn 455

RESULT 4  
AAU29061  
ID AAU29061 standard; Protein; 515 AA.  
XX  
AC AAU29061;  
XX  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Human PRO polypeptide sequence #38.  
XX  
KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;  
KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;  
KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;  
KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.  
XX  
OS Homo sapiens.

XX WO200168848-A2.  
XX 20-SEP-2001.  
XX  
XX 28-FEB-2001; 2001WO-US06520.  
XX  
XX 01-MAR-2000; 2000WO-US05601.  
XX 02-MAR-2000; 2000WO-US05841.  
XX 03-MAR-2000; 2000US-187202P.  
XX 06-MAR-2000; 2000US-186968P.  
XX 14-MAR-2000; 2000US-189320P.  
XX 14-MAR-2000; 2000US-189328P.  
XX 15-MAR-2000; 2000WO-US06884.  
XX 21-MAR-2000; 2000US-190828P.  
XX 21-MAR-2000; 2000US-191007P.  
XX 21-MAR-2000; 2000US-191048P.  
XX 21-MAR-2000; 2000US-191314P.  
XX 28-MAR-2000; 2000US-192553P.  
XX 29-MAR-2000; 2000US-193032P.  
XX 29-MAR-2000; 2000US-193053P.  
XX 30-MAR-2000; 2000WO-US08439.  
XX 04-APR-2000; 2000US-194449P.  
XX 04-APR-2000; 2000US-194647P.  
XX 11-APR-2000; 2000US-195975P.  
XX 11-APR-2000; 2000US-196000P.  
XX 11-APR-2000; 2000US-196187P.  
XX 11-APR-2000; 2000US-196690P.  
XX 18-APR-2000; 2000US-196820P.  
XX 18-APR-2000; 2000US-198121P.  
XX 18-APR-2000; 2000US-198585P.  
XX 25-APR-2000; 2000US-199397P.  
XX 25-APR-2000; 2000US-199550P.  
XX 25-APR-2000; 2000US-199654P.  
XX 03-MAY-2000; 2000US-201516P.  
XX 17-MAY-2000; 2000WO-US13705.  
XX 22-MAY-2000; 2000WO-US14042.  
XX 30-MAY-2000; 2000WO-US14941.  
XX 02-JUN-2000; 2000WO-US15264.  
XX 05-JUN-2000; 2000US-209832P.  
XX 28-JUL-2000; 2000WO-US20710.  
XX 22-AUG-2000; 2000US-0644848.  
XX 24-AUG-2000; 2000WO-US23328.  
XX 08-NOV-2000; 2000WO-US30952.  
XX 01-DEC-2000; 2000WO-US32678.  
XX 20-DEC-2000; 2000WO-US34956.  
XX  
XX (GETH ) GENENTECH INC.  
XX  
XX Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;  
XX Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;  
XX  
XX WPI; 2001-602746/68.  
XX N-PSDB; AAS45962.  
XX  
XX Novel nucleic acids encoding PRO polypeptides, used to diagnose the  
XX presence of tumours, such as prostate and breast tumours, in mammals and  
XX to screen for modulators of the compounds -  
XX  
XX Claim 11; Fig 76; 774pp; English.  
XX  
XX Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.  
XX The PRO polypeptides and their associated nucleic acids can be used to  
XX detect the presence of a tumour in a mammal by comparing the level of  
XX expression of a PRO polypeptide in a test sample of cells from the animal  
XX and a control sample of normal cells, whereby a higher level of  
XX expression in the test sample indicates the presence of a tumour in the  
XX mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats  
XX and rabbits but are preferably human. The polypeptides can be used to  
XX stimulate tumour necrosis factor (TNF) alpha release from human blood,  
XX when contacted with it. A specific polypeptide can be used to stimulate  
XX the proliferation or differentiation of chondrocyte cells. The PRO  
XX proteins can be used to determine the presence of tumours and also

CC susceptibility to tumour development, particularly adrenal, lung, colon,  
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian  
CC subjects. The oligonucleotide probes specific for the pro nucleic acids  
CC can be used for genetic analysis of individuals with genetic disorders.  
XX  
XX Sequence 515 AA;

Query Match 76.6%; Score 2307.5; DB 22; Length 515;  
Best Local Similarity 93.7%; Pred. No. 1.5e-210;  
Matches 434; Conservative 3; Mismatches 15; Indels 11; Gaps 4;  
QY 1 MAPRGCAHPPPPSPQACVCGKMLAGFWILCLTYGYSWGQALEEEEGALLA 60  
Db 1 maprgcaghppppspqacvcpqkmlangalagfwilclltygyswgaleeeegalla 60  
QY 61 QAGEKLEPSTTSQPHLIFILADDQFRDVGHGSEIKTPTLDKLAAGVKLENYVQP 120  
Db 61 qageklepsttsqphliladdqgfrdvgyhgseiktptldklaegvklennyvqp 120  
QY 121 ICTPSRSQFITGKYQIHTGLQHSIIRTPQNCPLDNLTPQKLKEGYSTHMGKWHLG 180  
Db 121 ictpsrsqfitgkyqihgtglqhsiiirtpqncpldnltpqklkevgysthmgvkwkwhlg 180  
QY 181 FYRKECMPTRRGFTFFGSLGSDYTHYKCDSPGCMGYDLYENDNAWDYDNGIYSTQ 240  
Db 181 fnrkecmptrrgfdtffgslgsgdythykcdspgcmgydlyendnaawdydngiystq 240  
QY 241 MYTORVQOILASHNPTKPIFLYTAQVHSPLOAPGRYFEHYRSIIININRRRYAAMLSC 300  
Db 241 mytrqvqilashnptkpiflytaqavhsploapgrgyfeyhrysiinirrryaamlsc 300  
QY 301 DEAINNVTLKTYGYFNNSIIYSSDNGGQPTAGGSNWPLRGSKGYWEGGIRAVGFVH 360  
Db 301 deainnvtlalktygyfnnsiiyssdnggqptaggsnwnplrgskgyweegiravgfvh 360  
QY 361 SPLKNGKGVCKEVPVHTIDWYPTLISLAEGQIDEDIQDGYDIWETISEGLRSPRVDLH 420  
Db 361 splkngktvckelvhidwypthlislaegqidediqldgydiwetiseglrsprvdlh 420  
QY 421 NIDPIYT--KAKNGSWAAGVGIIWNTAIQSAIRVQHWKLLTGN 460  
Db 421 nidp-ycprkmpgqqamsg-----tlqss---qpsecstgn 455

RESULT 5  
AAB85482  
ID AAB85482 standard; Protein; 569 AA.  
XX  
AC AAB85482;  
DB  
DT 25-SEP-2001 (first entry)  
XX  
DE Human 25278 sulfatase polypeptide.  
XX  
XX Sulfatase; 25278 sulfatase; nootropic; neuroprotective; antibacterial;  
KW antiinflammatory; vasotropic; antitumor; gene therapy; human.  
XX  
OS Homo sapiens.  
XX  
XX WO200155411-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 31-JAN-2001; 2001WO-US03266.  
XX  
PR 31-JAN-2000; 2000US-0495823.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Glucksmann MA, Williamson M, Rudolph-Owen LA, Tsai F;  
XX  
DR WPI; 2001-476214/51.

DR N-PSDB; AAH46862, AAH46866.  
XX  
XX Novel human sulfatase polypeptides useful for treating and diagnosing  
PT sulfatase-related disorders such as cerebrovascular diseases, acute  
PT meningitis, multiple sclerosis, degenerative diseases and tumor -  
XX  
XX Claim 9; Fig 10; 180pp; English.  
XX  
CC The invention provides 22438, 23553, 25278 or 26212 human sulfatase  
CC polypeptides and polynucleotides. The sulfatase genes and polypeptides  
CC are useful for treating disorders involving the brain such as  
CC cerebrovascular diseases, infections such as acute meningitis,  
CC demyelinating diseases including multiple sclerosis, degenerative  
CC diseases affecting the cerebral cortex including Alzheimer's disease  
CC and Pick disease, spinocerebellar degenerations including spinocerebellar  
CC ataxias including Friedreich ataxia, and ataxia telangiectasia,  
CC degenerative diseases affecting motor neurons including amyotrophic  
CC lateral sclerosis, inborn errors of metabolism such as leukodystrophies,  
CC toxic and acquired metabolic diseases, including vitamin deficiencies,  
CC and neurocutaneous syndromes (phakamatoses) including neurofibromatosis.  
CC The present sequence represents a human 25278 sulfatase polypeptide.  
XX  
XX Sequence 569 AA;  
QY 27 MGALAGFWILCLTYGYSWGQALEEEEGALLA----QAGEKLEPSTTSQPHLIFIL 82  
Db 1 mhtltisfslvslsfyglswdwa----kpsfvadpggeage--qpsaapqpqhifil 53  
QY 83 ADDGFRDVGHGSEIKTPTLDKLAAGVKLENYVQPICTPSRSQFITGKYQIHTGLQH 142  
Db 54 tddggyhadvghsgdiectptldriaakgklyenyvdiqictpsrsqiltgyqihgtglq 113  
QY 143 SIIRPTQNCPLDNLTPQKLKEGYSTHMGKWHLGFKYRKECMPTRRGFTFFGSLG 202  
Db 114 siirpqpnclpldqvtlpgklqegysthmvgkwhlgfyrkeclptrrgfdtflgsltg 173  
QY 203 SGDYTHYKCDSPGCMGYDLYENDNAWDYDNGIYSTQMYTORVQOILASHNPTKPIFLY 262  
Db 174 nvdytydncdpgvcgfdihgenvawgl-sgdystmlyaqrashilashspqrplfly 232  
QY 263 IAYQAVHSPLOAPGRYFEHYRSIIININRRRYAAMLSCIDEAINNVTLKTYGYFNNSII 322  
Db 233 vafqavhtplqspreylyryrtmgnvarrryaamvmtcmdeavrnitwalkfygfnnsvi 292  
QY 323 IYSSDNGGQPTAGGSNWPLRGSKGYWEGGIRAVGFVHSPLLKNGKGVCKEVPVHTIDWYP 382  
Db 293 ifssdnggqptfsggsnwnplrgktywegvrglgfwhspilkrkqtrtsalmhitdwp 352  
QY 383 TLISLAEGQIDEDIQDGYDIWETISEGLRSPRVDLIHNDPIYTKAKNGSWAAGYGIWN 442  
Db 353 tlvglagtttsaadlgdydwpaalsegrasprteihlnidplynhadglsleggfign 412  
QY 443 TAIQSAIRVQHWKLLTGNPGYSDMWVPPQSFNSLGNPRWHNERITSSSTCKSVLWLENITADP 502  
Db 413 tavqaairvgewklltgdpgygdwlpptqtlatf-pgswnlnlernasvrqavwlfnlsadp 471  
QY 503 YERVDLSNRPYIGIVKLLRLRLSQFNKTAVPVRYPPKDPSPRNPLNGG 549  
Db 472 yeredlagqrpdvvrtllarlaeynrtaiprvypaenprahpfdngg 518  
RESULT 6  
AAU14421  
ID AAU14421 standard; Protein; 289 AA.  
XX  
AC AAU14421;  
XX  
DT 24-OCT-2001 (first entry)

XX DE Human novel protein #292.  
XX KW Human; novel protein; Antianemic; osteopathic; antiinflammatory;  
XX KW immunomodulatory; cytostatic; neuroprotective; vulnery; nootropic;  
XX KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;  
XX KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;  
XX KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;  
XX KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;  
XX KW tissue regeneration; immune disorder.  
XX OS Homo sapiens.  
XX KW WO200155437-A2.  
XX PN 02-AUG-2001.  
XX PD 25-JAN-2001; 2001WO-US02623.  
XX PF 25-JAN-2000; 2000US-0491404.  
XX PR (HYSE-) HYSEQ INC.  
XX PA Tang YT, Liu C, Drmanac RT;  
XX FI WPI; 2001-451939/48.  
XX DR N-PSDB; AAS22726.  
XX DR Isolated polypeptides useful for treating anti-inflammatory diseases,  
XX PT nervous system disorders, and for regenerating bone and cartilage -  
XX PS Example 4; Page 809-810; 894pp; English.  
XX CC The invention relates to polynucleotides encoding novel human  
XX CC proteins or their active domains. The polypeptides, polynucleotides and  
XX CC antibodies raised against the polypeptides are used in a method of  
XX CC treatment of a mammal and prevention of disorders caused by the aberrant  
XX CC protein expression or activity. The polypeptides can be used as  
XX CC molecular weight markers, food supplements, and in antibody production.  
XX CC The polypeptides are used to identify compounds which bind to the  
XX CC polypeptides. Polynucleotides of the invention are used as probes and  
XX CC primers, for sequencing, for chromosome or gene mapping, in the  
XX CC production of recombinant proteins, and in generating anti-sense DNA or  
XX CC RNA and in gene therapy. Polypeptides of the invention can be used to  
XX CC target drugs to a tumour, in assays to determine biological activity, to  
XX CC raise antibodies/elicit an immune response, to determine quantitative  
XX CC protein levels, as tissue markers, and to isolate receptors or ligands.  
XX CC Polypeptides of the invention may also be useful in treating platelet  
XX CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,  
XX CC ligament and/or nerve tissue, wound healing, treating burns, promoting  
XX CC the proliferation, differentiation and survival of stem cells, as a  
XX CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,  
XX CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral  
XX CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or  
XX CC fungal infection or from autoimmunity, cancer, allergy, asthma,  
XX CC graft-versus-host disease, eczema, haemophilia, thrombosis,  
XX CC anti-inflammatory diseases, nervous system disorders, and infection.  
XX CC The present sequence represents a protein of the invention.  
XX SQ Sequence 289 AA;

Query Match 51.9%; Score 1562; DB 22; Length 289;  
Best Local Similarity 100.0%; Pred. No. 5.8e-140;  
Matches 289; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 24 MIANGALAGFWILCLLYGYLSWQALBEEBEGALLAQAGEKLEPSTSTSQPHLIFILA 83  
DB 1 mlANGalagfwilclltygylswgaleeeegallagakelepststsqphliffila 60  
QY 84 DDQCFRDVGYHSGSEIKPTLDKLAAGVKLENYYVQICTPSRSQFTGKYIHTGLQHS 143  
DB 61 ddqgfrdvgyhsgseikptldklaagvklennyvqpictpsrsqftgkyihtgqlghs 120

QY 144 IIRPTQPNCLPLDNATLPQKLKEVGYSTHVMGVKWHLGEYRKECMPTRRGFTDFGSLGGS 203  
DB 121 IIRPTQPNCLPLDNATLPQKLKEVGYSTHVMGVKWHLGEYRKECMPTRRGFTDFGSLGGS 180  
QY 204 GDYTHYKCDSPGCGYDLYENDNAANDYNGIYSTQMYTORVQOILASHNPTKPIFYI 263  
DB 181 gdythykcdspgcmgydlyendnaawdynglystqmytqrvgqilashnptkpiilyi 240  
QY 264 AYQAVHSPLOQPGYFVFEHYRSTIINRRRYAAMLSCDLDEAINNVTLALK 312  
DB 241 ayqavhsploqpgryfeyhrsliinrrrryaamisclddeainnvtlalk 289  
RESULT 7  
AAB51184  
ID AAB51184 standard; Protein; 533 AA.  
XX AC AAB51184;  
XX DT 21-MAR-2001 (first entry)  
XX DE Human sulfatase protein B SEQ ID NO:13.  
XX KW Human; iduronate 2-sulfatase; IDS; glycosylated; gene therapy;  
KW mucopolysaccharidosis inhibitor; IDS deficiency disorder;  
KW Hunter syndrome; mucopolysaccharidosis type II.  
XX OS Homo sapiens.  
XX PN US6153188-A.  
XX PD 28-NOV-2000.  
XX PF 12-FEB-1999; 99US-0249003.  
XX PR 17-DEC-1992; 92US-0991973.  
XX PR 28-NOV-1994; 94US-0345212.  
XX PR 12-NOV-1991; 91US-0790362.  
XX PA (WOMEN-) WOMEN'S & CHILDREN'S HOSPITAL.  
XX PI Wilson PJ, Anson DS, Occhiodoro T, Bielicki J, Clements PR;  
PI Hopwood JJ, Morris CP;  
XX DR WPI; 2001-060076/07.  
XX PT New highly glycosylated recombinant human iduronate 2-sulfatase (IDS)  
PT useful for diagnosing or treating subjects suspected of having or  
PT suffering from IDS deficiency disorders, e.g. Hunter syndrome  
PT (mucopolysaccharidosis-II) -  
XX PS Example 1; Fig 3; 53pp; English.  
XX CC The present invention describes a recombinant human iduronate 2-sulfatase  
CC (IDS). The recombinant IDS is more highly glycosylated than the naturally  
CC occurring enzyme isolated from human tissue. The recombinant human IDS  
CC can be produced in Chinese Hamster Ovary (CHO) cells or in a human cell.  
CC The recombinant IDS comprises a fusion protein. It is a  
CC mucopolysaccharidosis inhibitor and can be used in gene therapy.  
CC The recombinant IDS is useful in treating and diagnosing subjects  
CC suffering from or suspected of having IDS deficiency disorders, e.g.  
CC Hunter syndrome (mucopolysaccharidosis type II). The present sequence  
CC represents a human sulfatase which is used in comparison with the  
CC IDS sequence in an example from the present invention.  
XX SQ Sequence 533 AA;

Query Match 47.0%; Score 1415.5; DB 22; Length 533;  
Best Local Similarity 49.1%; Pred. No. 1.4e-125;  
Matches 278; Conservative 80; Mismatches 131; Indels 57; Gaps 8;











PR	17-NOV-2000;	2000US-0249300.	
PR	01-DEC-2000;	2000US-0250160.	
PR	01-DEC-2000;	2000US-0250391.	
PR	05-DEC-2000;	2000US-0251030.	
PR	05-DEC-2000;	2000US-0251988.	
PR	05-DEC-2000;	2000US-0256719.	
PR	06-DEC-2000;	2000US-0251479.	
PR	08-DEC-2000;	2000US-0251856.	
PR	08-DEC-2000;	2000US-0251868.	
PR	08-DEC-2000;	2000US-0251869.	
PR	08-DEC-2000;	2000US-0251989.	
PR	08-DEC-2000;	2000US-0251990.	
PR	11-DEC-2000;	2000US-0254097.	
PR	05-JAN-2001;	2001US-0259678.	
XX			
XX	(HUMA-)	HUMAN GENOME SCI INC.	
XX			
PI	Rosen CA, Barash SC, Ruben SM;		
XX			
XX	WPI; 2001-465566/50.		
XX	N-PSDB; AAS41205.		
XX			
PT	Novel polypeptides and polynucleotides useful for diagnosing,		
PT	preventing, treating neural, immune system, muscular, reproductive,		
PT	pulmonary, cardiovascular, renal, proliferative disorders and cancerous		
PT	diseases		
XX			
PS	Claim 11; SEQ ID No 1331; 1180pp; English.		
XX			
CC	The present invention relates to the isolation of novel human enzyme		
CC	polypeptides, and the cDNA (AA540785-AAS41684) and genomic sequences		
CC	encoding them. The enzyme polypeptides of the invention may comprise the		
CC	functional classes of oxidoreductases, transferases, hydrolases, lyases,		
CC	isomerases or ligases. The sequences of the invention are useful in the		
CC	diagnosis, treatment, prevention and/or prognosis of a wide range of		
CC	disorders including hyperproliferative disorders (e.g. cancer),		
CC	immunodeficiency disorders (e.g. AIDS) autoimmune disorders		
CC	(e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),		
CC	metabolic disorders (e.g. phenylketonuria), inflammatory disorders		
CC	(e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),		
CC	blood-related disorders (e.g. haemophilia), reproductive disorders		
CC	(e.g. infertility) and infectious disorders (e.g. influenza). The		
CC	polynucleotides of the invention can also be used in gene therapy.		
CC	AAU22915-AAU23814 represent the novel human enzyme polypeptides of the		
CC	invention.		
CC	Note: The sequence data for this patent did not form part of the printed		
CC	specification, but was obtained in electronic format directly from WIPO		
CC	at ftp.wipo.int/pub/published_pct_sequences.		
XX			
SQ	Sequence 128 AA;		
	Query Match 20.0%; Score 603.5; DB 22; Length 128;		
	Best Local Similarity 94.2%; Pred. No. 2.9e-49;		
	Matches 113; Conservative 0; Mismatches 6; Indels 1; Gaps		
QY	90 DVGVHGSEIKTPTLDKLAAGVKLENVYVQICTPSRSQFITCKYQIHTGLQHSIIIRPTQ 149		
Db	6 dvgyhgseiktptldkllaegvklennyvqictpsrsqfitgkyqlhgqlnsirpq 65		
QY	150 PNCPLDNLATLPQKLKEVGYSTHMVGKWHUGFYRKCCMPTRRGFDTFFGSLLGS-GDYTT 208		
Db	66 pncpldnltpqklkevgysthmvgkwhlgyfrykccmptrrgfdtffgfpfokwglylt 125		
RESULT 14			
ABB63259			
ID	ABB63259 standard; Protein; 486 AA.		
XX			
AC	ABB63259;		
XX			
DT	26-MAR-2002 (first entry)		



Db 72 plsspsraglltgmpfrtgr-swipsgkdvalgrneltianlllkaggydtammgklhl 130  
QY 180 ---GFYRKECMPTRRGEDTFFGSLGSGDYTHYKCDSPGCGYDLYENDNAAWDYDNGI 236  
Db 131 naggdrtddpqaqmgfdy---slantagfvtdatld-----nakerprygmvyptgw 180  
QY 237 YSTOMYTORVQOI-----LASHNPTKPIFLYIAYQAVHSPLQAPGRYFEHYRS 284  
Db 181 lrngqptpradkmsgeyvssevvnwldnkkdkpfflyavaftehshplaspkkyldmysq 240  
QY 285 IININRRR-----YAAWLSCLDEAINNVTLALATYGFYNNNSIIYS 325  
Db 241 ymseayqkqhpdlfygdwadkpwrgvgeyyanisylgaqvkvldkikamgeedhtivift 300  
QY 326 SDNGG-----QPTAGGSNNWPLRGSKGTWEGGIRAVGFV-----HSPLLKNKGTVCCK 372  
Db 301 sdngpvtrearkvyelnlagetdglgrkdnlweggirvpaiikygkhlp----qgmvsd 356  
QY 373 EPVHITDWYPTLISLAEGQIDEDQLDGYDIWETISEGLRSPRVDILHNIDPIYTKAKNG 432  
Db 357 tpvygldwmptlakmmnfkiptdrtdfgeslvpvledkalkrekplifgidmpfqddptd 416  
QY 433 SWAAGYGIWNTAIQSAIRVOHWKLLTGNPGYSDWVPQSFNSNLGNRWHNERITSSTGKS 492  
Db 417 ew-----alrdgdwkm-----iidrnknkp 435  
QY 493 VWLFNITADPYERVDLSNRYPGIVKLLRLRSQFNKTAV 531  
Db 436 kylvnlksdryetlnlllgkxpdiiekqmygkflky-ktidi 473

Search completed: July 23, 2002, 19:02:50  
Job time: 6936 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 19:04:14 ; Search time 22.71 seconds  
(without alignments)  
2327.131 Million cell updates/sec

Title: US-09-495-823-7

Perfect score: 550

Sequence: 1 MAPRGACGHPPPSPQACVC.....VPVRYPPKDRSPRLNGGV 550

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

PIR\_71:.\*  
1: piri:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	11	2.0	473	2 I54210	N-acetylgalactosam
2	10	1.8	289	2 AI0374	probable thioredox
3	10	1.8	533	1 KJHUB	N-acetylgalactosam
4	10	1.8	535	1 A44475	N-acetylgalactosam
5	8	1.5	176	2 S76001	phosphoribosylamin
6	8	1.5	219	2 A23385	two-component resp
7	8	1.5	330	2 A83230	3-oxoacyl-[acyl-ca
8	8	1.5	381	2 AD1113	hypothetical prote
9	8	1.5	381	2 AH1474	hypothetical prote
10	8	1.5	452	2 T29618	hypothetical prote
11	8	1.5	551	2 S07089	arylsulfatase (EC
12	8	1.5	551	2 S01793	arylsulfatase (EC
13	8	1.5	567	2 A37362	arylsulfatase (EC
14	8	1.5	647	2 S61973	hypothetical prote
15	8	1.5	1216	2 AH1335	pyruvate-flavodoxi
16	8	1.5	1420	2 B57062	SRB9 protein - yea
17	7	1.3	82	2 AD2287	hypothetical prote
18	7	1.3	108	2 AG0271	translation initia
19	7	1.3	114	2 E70327	hypothetical prote
20	7	1.3	123	2 H72468	hypothetical prote
21	7	1.3	123	2 H84014	hypothetical prote
22	7	1.3	155	2 AI2450	hypothetical prote
23	7	1.3	159	2 AH1524	hypothetical prote
24	7	1.3	180	2 C39415	fimbrial protein s
25	7	1.3	198	2 T45541	septum-promoting G
26	7	1.3	200	2 T49590	hypothetical prote
27	7	1.3	220	2 C83292	probable glutathio
28	7	1.3	229	2 F84685	hypothetical prote
29	7	1.3	232	2 AH0252	conserved hypothet

30	7	1.3	237	2 AC0758	precorrin-2 C20-me
31	7	1.3	238	1 Q0BE74	US4 protein - huma
32	7	1.3	241	2 T20432	hypothetical prote
33	7	1.3	248	2 AI2175	hypothetical prote
34	7	1.3	277	2 AD3348	creatininase (EC 3
35	7	1.3	278	2 G81148	pantoate-beta-ala
36	7	1.3	286	2 E91262	probable aldolase
37	7	1.3	286	2 H86102	probable aldolase
38	7	1.3	286	2 AI2390	hypothetical prote
39	7	1.3	294	2 AB0115	probable carbon-ni
40	7	1.3	298	2 JC7568	kidney inhibitor o
41	7	1.3	301	2 AF2004	33kD chaperonin, h
42	7	1.3	302	2 S75112	hypothetical prote
43	7	1.3	303	2 T15835	hypothetical prote
44	7	1.3	304	2 B83951	tRNA pseudouridine
45	7	1.3	319	2 A98140	glycosyl transfera

ALIGNMENTS

RESULT 1

I54210

N-acetylgalactosamine-4-sulfatase (EC 3.1.6.12) precursor - rat (fragment)  
N;Alternate names: arylsulfatase (EC 3.1.6.1) B [misidentification]  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 02-Jun-2000  
C;Accession: I54210

R;Kunieda, T.

Genomics 29, 582-587, 1995

A;Title: Mucopolysaccharidosis type VI in rats: Isolation of cDNAs encoding arylsulfatase  
A;Reference number: I54210; MUID:96121368

A;Accession: I54210

A;Status: preliminary; translated from GB/EMBL/DBDJB

A;Molecule type: mRNA

A;Residues: 1-473 <RES>

A;Cross-references: GB:D49434; NID:g1065603; PIDN:BAA08412.1; PID:g1089794

C;Genetics:

A;Gene: ARSB

C;Superfamily: animal sulfatase

C;Keywords: sulfuric ester hydrolase

F;31/Modified site: 3-oxoalanine (Cys) #status predicted

Query Match 2.0% Score 11; DB 2; Length 473;  
Best Local Similarity 100.0%; Pred. No. 0.012;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 170 STHMVGKWHLG 180

|||||

Db 79 STHMVGKWHLG 89

RESULT 2

AI0374

probable thioredoxin YP03082 [imported] - Yersinia pestis (strain CO92)

C;Species: Yersinia pestis

C;Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001

C;Accession: AI0374

R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G  
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrel  
Nature 413, 523-527, 2001

A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A;Reference number: AB0001; MUID:21470413; PMID:11586360

A;Accession: AI0374

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-289 <KUR>

A;Cross-references: GB:AL590842; PIDN:CAC92324.1; PID:g15981035; GSPDB:GN00175

C;Genetics:

A;Gene: YP03082

Query Match 1.8%; Score 10; DB 2; Length 289;  
Best Local Similarity 100.0%; Pred. No. 0.085;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 100 TPTLDKLAEE 109  
|||||  
Db 42 TPTLDKLAEE 51

RESULT 3  
KJH0AB  
N-acetylglactosamine-4-sulfatase (EC 3.1.6.12) precursor [validated] - human  
N:Alternate names: arylsulfatase B (ASB); chondroitinase; chondroitinsulfatase; G4S; N-acetylglactosaminidase; arylsulfatase B (ASB); chondroitinase; chondroitinsulfatase; G4S;  
C:Species: Homo sapiens (man)  
C:Date: 31-Dec-1993 #sequence\_revision 27-Oct-1995 #text\_change 08-Dec-2000  
C:Accession: S35990; S33307; A35078; A45659; A42449; B42449; C42449; I34217; A56865  
R:Peters, C.W.B.  
submitted to the EMBL Data Library, March 1993  
A:Reference number: S35990  
A:Accession: S35990  
A:Molecule type: DNA  
A:Residues: 1-533 <PET>  
A:Cross-references: EMBL:X72735; NID:g289009; PIDN:CAA51272.1; PID:g825628  
R:Modaressi, S.; Rupp, K.; von Figura, K.; Peters, C.  
Biol. Chem. Hoppe-Seyler 374, 327-335, 1993  
A:Title: Structure of the human arylsulfatase B gene.  
A:Reference number: S33307; MUID:93332648  
A:Accession: S33307  
A:Molecule type: DNA  
A:Residues: 1-104 <MOD>  
A:Cross-references: EMBL:X72735; EMBL:X72736; EMBL:X72737; EMBL:X72738; EMBL:X72739; EMBL:X72740  
A:Note: The enzyme is referred to as EC 3.1.6.9  
R:Peters, C.; Schmidt, B.; Rommerskirch, W.; Rupp, K.; Zuehlendorf, M.; Vingron, M.; Meyer, J. Biol. Chem. 265, 3374-3381, 1990  
A:Title: Phylogenetic conservation of arylsulfatases. cDNA cloning and expression of human arylsulfatase B.  
A:Reference number: A35078; MUID:90153994  
A:Accession: A35078  
A:Molecule type: mRNA  
A:Residues: 1-357, 'V', 359-533 <PE>  
A:Cross-references: GB:052225; NID:g179076; PIDN:AAA51784.1; PID:g179077  
A:Note: parts of this sequence were determined by protein sequencing  
A:Note: the enzyme is referred to as EC 3.1.6.1  
R:Litjens, T.; Morris, C.P.; Gibson, G.J.; Beckmann, K.R.; Hopwood, J.J.  
Biochem. Int. 24, 209-215, 1991  
A:Title: Human N-acetylglactosamine-4-sulphatase: protein maturation and isolation of cDNA.  
A:Reference number: A45659; MUID:92028992  
A:Accession: A45659  
A:Molecule type: DNA; protein  
A:Residues: 1-104 <LIT>  
A:Cross-references: GB:S57777; NID:g236697; PIDN:AAB19988.1; PID:g236698  
A:Note: sequence extracted from NCBI backbone (NCBIN:57777, NCBIP:57776)  
A:Note: the enzyme is referred to as EC 3.1.6.1  
A:Note: parts of this sequence, including the amino end of the mature protein, were determined by protein sequencing  
A:Note: a form is described with a proteolytic cleavage somewhere between residue 450 and 460  
R:Jin, W.D.; Jackson, C.E.; Desnick, R.J.; Schuchman, E.H.  
Am. J. Hum. Genet. 50, 795-800, 1992  
A:Title: Mucopolysaccharidosis type VI: identification of three mutations in the arylsulfatase B gene.  
A:Reference number: A42449; MUID:92197625  
A:Accession: A42449  
A:Molecule type: mRNA  
A:Residues: 115-116, 'R', 118 <JIN>  
A:Cross-references: GB:S90729; NID:g247486; PIDN:AAB21831.1; PID:g247487  
A:Note: sequence extracted from NCBI backbone (NCBIN:90729, NCBIP:90731)  
A:Accession: B42449  
A:Molecule type: mRNA  
A:Residues: 234-235, 'P', 237-238 <J2>  
A:Cross-references: GB:S90736; NID:g247488; PIDN:AAB21832.1; PID:g247489  
A:Note: sequence extracted from NCBI backbone (NCBIN:90736, NCBIP:90739)  
A:Accession: C42449  
A:Molecule type: mRNA  
A:Residues: 403-404, 'Y', 406-407 <J13>

A:Cross-references: GB:S90743; NID:g247490; PIDN:AAB21833.1; PID:g247491  
A:Note: sequence extracted from NCBI backbone (NCBIN:90743, NCBIP:90747)  
A:Note: the enzyme is referred to as EC 3.1.6.1  
A:Note: these mutations give rise to mucopolysaccharidosis type VI, Maroteaux-Lamy di-  
R:Schmidt, B.; Selmer, T.; Ingendoh, A.; von Figura, K.  
Cell 82, 271-278, 1995  
A:Title: A novel amino acid modification in sulfatases that is defective in multiple  
A:Reference number: A57113; MUID:95354208  
A:Contents: annotation; identification of 3-oxoalanine, 2-amino-3-oxopropanoic acid  
R:Schuchman, E.H.; Jackson, C.E.; Desnick, R.J.  
Genomics 6, 149-158, 1990  
A:Title: Human arylsulfatase B: MOPAC cloning, nucleotide sequence of a full-length cDNA.  
A:Reference number: I54217; MUID:90152677  
A:Accession: I54217  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-357, 'V', 359-375, 'M', 377-533 <RES>  
A:Cross-references: GB:M32373; NID:g179029; PIDN:AAA51779.1; PID:g179030  
R:Kobayashi, T.; Honke, K.; Jin, T.; Gasa, S.; Miyazaki, T.; Makita, A.  
Biochim. Biophys. Acta 1159, 243-247, 1992  
A:Title: Components and proteolytic processing sites of arylsulfatase B from human pl  
A:Reference number: A56865; MUID:93003385  
A:Accession: A56865  
A:Molecule type: protein  
A:Residues: 41-55; 424-425, 'X', 427-454; 466-483 <ROB>  
A:Experimental source: Placenta  
A:Note: sequence modified after extraction from NCBI backbone  
A:Note: the fragments shown are the amino ends of the alpha, gamma, and beta chains o  
C:Comment: This enzyme is frequently misidentified as EC 3.1.6.1.  
C:Genetics:  
A:Gene: GDB:ARSB  
A:Cross-references: GDB:I19008; OMIM:253200  
A:Map position: 5q11-5q13  
A:Introns: 104/3; 167/1; 230/3; 300/1; 381/2; 405/1; 446/1  
A:Note: defects in this gene can cause mucopolysaccharidosis type VI, Maroteaux-Lamy  
C:Function:  
A:Description: hydrolyzes N-acetylglactosamine-4-sulfate units in chondroitin sulfate  
C:Superfamily: animal sulfatase  
C:Keywords: glycoprotein; lysosomal storage disease; lysosome; sulfuric ester hydrolase  
F:1-40/Domain: signal sequence #status predicted <SIG>  
F:41-423/Product: alpha chain #status predicted <AMAT>  
F:424-465/Product: gamma chain #status predicted <GMAT>  
F:466-533/Product: beta chain #status experimental <BMAT>  
F:91/Modified site: 3-oxoalanine (Cys) #status experimental  
F:188,279,366,458/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:291/Binding site: carbohydrate (Asn) (covalent) #status absent  
F:426/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 1.8%; Score 10; DB 1; Length 533;  
Best Local Similarity 100.0%; Pred. No. 0.15;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 THMVGKWHLG 180  
|||||  
Db 140 THMVGKWHLG 149

RESULT 4  
A44475  
N-acetylglactosamine-4-sulfatase (EC 3.1.6.12) precursor - cat  
N:Alternate names: arylsulfatase B (ARSB); chondroitinase; chondroitinsulfatase; G4S;  
C:Species: Felis silvestris catus (domestic cat)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C:Accession: A44475  
R:Jackson, C.E.; Yuhki, N.; Desnick, R.J.; Haskins, M.E.; O'Brien, S.J.; Schuchman, E.  
Genomics 14, 403-411, 1992  
A:Title: Feline arylsulfatase B (ARSB): isolation and expression of the cDNA, compari  
A:Reference number: A44475; MUID:93052342  
A:Accession: A44475  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-535 <JAC>

A;Cross-references: GB:S48472; NID:g258855; PIDN:AAB23941.1; PID:g258856  
A;Experimental source: liver  
A;Note: sequence extracted from NCBI backbone (NCBIP:117976)  
C;Comment: This enzyme is frequently misidentified as EC 3.1.6.1.  
C;Function:  
A;Description: hydrolyzes N-acetylgalactosamine-4-sulfate units in chondroitin sulfate A  
C;Superfamily: animal sulfatase  
C;Keywords: glycoprotein; lysosomal storage disease; lysosome; sulfuric ester hydrolase  
F;1-42/Domain: signal sequence #status predicted <SIG>  
F;43-425/Product: alpha chain #status predicted <AMAT>  
F;426-467/Product: gamma chain #status predicted <GMAT>  
F;468-535/Product: beta chain #status predicted <BMAT>  
F;93/Modified site: 3-oxoalanine (Cys) #status predicted  
F;190,281,428,460/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 1.8%; Score 10; DB 1; Length 535;  
Best Local Similarity 100.0%; Pred. No. 0.15;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 THMVGKWHLG 180  
|||||  
Db 142 THMVGKWHLG 151

RESULT 5  
S76001  
phosphoribosylaminoimidazole carboxylase (EC 4.1.1.21) catalytic chain - Synecocystis s  
N;Alternate names: protein sll0901  
C;Species: Synecocystis sp.  
A;Variety: PCC 6803  
C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
C;Accession: S76001  
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis  
s.  
A;Reference number: S74322; MUID:97061201  
A;Accession: S76001  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-176 <KAN>  
A;Cross-references: EMBL:D64006; GB:AB001339; NID:g1001291; PIDN:BAAL0848.1; PID:g100136  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C;Genetics:  
A;Gene: pure  
C;Superfamily: phosphoribosylaminoimidazole carboxylase catalytic chain; phosphoribosyla  
C;Keywords: carbon-carbon lyase; carboxy-lyase; purine nucleotide biosynthesis  
F;6-139/Domain: phosphoribosylaminoimidazole carboxylase catalytic chain homology <PCC>

Query Match 1.5%; Score 8; DB 2; Length 176;  
Best Local Similarity 100.0%; Pred. No. 6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 248 QILASHNP 255  
|||||  
Db 131 QILASHNP 138

RESULT 6  
AC2385  
two-component response regulator all4635 [imported] - Anabaena sp. (strain PCC 7120)  
C;Species: Anabaena sp.  
A;Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120  
C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002  
C;Accession: AC2385  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriquchi  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S  
DNA Res. 8, 205-213, 2001  
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AC2385  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-219 <KUR>  
A;Cross-references: GB:BA000019; PIDN:BAW76334.1; PID:g17133772; GSPDB:GN001179  
A;Experimental source: strain PCC 7120  
C;Genetics:  
A;Gene: all4635  
C;Superfamily: regulatory protein comA; response regulator homology

Query Match 1.5%; Score 8; DB 2; Length 219;  
Best Local Similarity 100.0%; Pred. No. 7.3;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 513 PGIVKKLL 520  
|||||  
Db 127 PGIVKKLL 134

RESULT 7  
AB3230  
3-oxoacyl-[acyl-carrier-protein] synthase (EC 2.3.1.41) III PA3333 [similarity] - Pse  
C;Species: Pseudomonas aeruginosa  
C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 02-Feb-2001  
C;Accession: AB3230  
R;Stover, C.K.; Pham, X.O.; Erwin, A.L.; Miziochuchi, S.D.; Warrenner, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L  
; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
A;Reference number: AB2950; MUID:20437337  
A;Accession: AB3230  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-330 <STO>  
A;Cross-references: GB:AE004755; GB:AE004091; NID:g9949456; PIDN:AAG06721.1; GSPDB:GN  
A;Experimental source: strain PA01  
C;Genetics:  
A;Gene: fabH2; PA3333  
C;Superfamily: 3-oxoacyl-[acyl-carrier-protein] synthase III  
C;Keywords: acyltransferase

Query Match 1.5%; Score 8; DB 2; Length 330;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 EEGGALLA 60  
|||||  
Db 171 EEGGALLA 178

RESULT 8  
AD1113  
hypothetical protein lmo0307 [imported] - Listeria monocytogenes (strain EGD-e)  
C;Species: Listeria monocytogenes  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C;Accession: AD1113  
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec  
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;  
Ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla  
A;Title: Comparative genomics of Listeria species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: AD1113  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-381 <GLA>  
A;Cross-references: GB:NC\_003210; PIDN:CAD00834.1; PID:g16409672; GSPDB:GN00177  
A;Experimental source: strain EGD-e

C;Genetics:  
A;Gene: lmo0307

Query Match 1.5%; Score 8; DB 2; Length 381;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 GFYNNII 322  
| | | | | | | |  
DB 181 GFYNNII 188

RESULT 9

AH1474  
hypothetical protein lin0335 [imported] - Listeria innocua (strain Clip11262)  
C;Species: Listeria innocua  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C;Accession: AH1474  
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mak, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of Listeria species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: AH1474  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-381 <GLA>  
A;Cross-references: GB:AL592022; PIDN:CAC95568.1; PID:g16412764; GSPDB:GN00178  
A;Experimental source: strain Clip11262  
C;Genetics:  
A;Gene: lin0335

Query Match 1.5%; Score 8; DB 2; Length 381;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 GFYNNII 322  
| | | | | | | |  
DB 181 GFYNNII 188

RESULT 10

T29618  
hypothetical protein D1014.1 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C;Accession: T29618  
R;Du, Z.; Leinbac, D.  
submitted to the EMBL Data Library, March 1996  
A;Description: The sequence of C. elegans cosmid D1014.  
A;Reference number: Z20652  
A;Accession: T29618  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-452 <DUZ>  
A;Cross-references: EMBL:U53180; PIDN:AAA96290.1; GSPDB:GN00023; CESP:D1014.1  
A;Experimental source: strain Bristol N2; clone D1014  
C;Genetics:  
A;Gene: CESP:D1014.1  
A;Map position: 5  
A;Introns: 256/3; 336/2; 376/3; 405/3

Query Match 1.5%; Score 8; DB 2; Length 452;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 MVGKWHLG 180

Db 134 MVGKWHLG 141  
| | | | | | | |

RESULT 11

S07089  
arylsulfatase (EC 3.1.6.1) - sea urchin (Lytechinus pictus)  
C;Species: Lytechinus pictus (painted urchin)  
C;Date: 22-Jan-1993 #sequence\_revision 22-Jan-1993 #text\_change 26-Aug-1999  
C;Accession: S07089  
R;Yamada, K.; Akasaka, K.; Shimada, H.  
Eur. J. Biochem. 186, 405-410, 1989  
A;Title: Structure of sea-urchin arylsulfatase gene.  
A;Reference number: S07089; MUID:90092130

A;Accession: S07089  
A;Molecule type: DNA  
A;Residues: 1-551 <YAM>  
A;Cross-references: EMBL:X16679  
C;Genetics:  
A;Gene: Ars  
A;Introns: 103/2; 160/1; 197/3; 272/1; 327/1  
C;Superfamily: animal sulfatase  
C;Keywords: sulfuric ester hydrolase  
F:100/Modified site: 3-oxoalanine (Cys) #status predicted

Query Match 1.5%; Score 8; DB 2; Length 551;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 MVGKWHLG 180  
| | | | | | | |  
DB 153 MVGKWHLG 160

RESULT 12

S01793  
arylsulfatase (EC 3.1.6.1) precursor - sea urchin (Hemicentrotus pulcherrimus)  
C;Species: Hemicentrotus pulcherrimus  
C;Date: 30-Sep-1989 #sequence\_revision 30-Sep-1991 #text\_change 26-Aug-1999  
C;Accession: S01793  
R;Sasaki, H.; Yamada, K.; Akasaka, K.; Kawasaki, H.; Suzuki, K.; Saito, A.; Sato, M.; Eur. J. Biochem. 177, 9-13, 1988  
A;Title: cDNA cloning, nucleotide sequence and expression of the gene for arylsulfatase  
A;Reference number: S01793; MUID:89030699  
A;Accession: S01793  
A;Molecule type: mRNA  
A;Residues: 1-551 <SAS>  
A;Cross-references: EMBL:X17015; MID:g9432; PID:g9433  
A;Note: part of this sequence was confirmed by protein sequencing  
A;Note: the authors translated the codon CAC for residue 61 as Asp  
C;Superfamily: animal sulfatase  
C;Keywords: sulfuric ester hydrolase  
F:1-20/Domain: signal sequence #status predicted <SIG>  
F:21-551/Product: arylsulfatase #status predicted <MAT>  
F:100/Modified site: 3-oxoalanine (Cys) #status predicted

Query Match 1.5%; Score 8; DB 2; Length 551;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 MVGKWHLG 180  
| | | | | | | |  
DB 153 MVGKWHLG 160

RESULT 13

A37362  
arylsulfatase (EC 3.1.6.1) precursor - sea urchin (Strongylocentrotus purpuratus)  
C;Species: Strongylocentrotus purpuratus (purple urchin)  
C;Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 24-Nov-1999  
C;Accession: A37362



R;Yang, Q.; Angerer, L.M.; Angerer, R.C.

Dev. Biol. 135, 53-65, 1989

A;Title: Structure and tissue-specific developmental expression of a sea urchin arylsulphatase

A;Reference number: A37362; MUID:89357267

A;Accession: A37362

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-567 <YAN>

A;Cross-references: GB:M28404; GB:M25815; NID:g161440; PID:g161441

C;Superfamily: animal sulfatase

C;Keywords: sulfuric ester hydrolase

F;115/Modified site: 3-oxoalanine (Cys) #status predicted

#### Query Match

1.5%; Score 8; DB 2; Length 567;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 MVGKWHLG 180

|||||||

Db 168 MVGKWHLG 175

#### RESULT 14

S61973

hypothetical protein YPL093w - yeast (*Saccharomyces cerevisiae*)

N;Alternate names: hypothetical protein LPGA15w

C;Species: *Saccharomyces cerevisiae*

C;Date: 10-Apr-1996 #sequence\_revision 19-Apr-1996 #text\_change 05-Nov-1999

C;Accession: S61973

R;Wang, Y.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Hall, J.; Storms, R.K.; Vo

submitted to the EMBL Data Library, December 1995

A;Description: The sequence of *Saccharomyces cerevisiae* chromosome XVI left arm.

A;Reference number: S61959

A;Accession: S61973

A;Molecule type: DNA

A;Residues: 1-647 <WAN>

A;Cross-references: EMBL:U43281; NID:g1151218; PID:g1151233; GSPDB:GN00016; MIPS:YPL093w

C;Genetics:

A;Gene: MIPS:YPL093w

A;Map position: 16L

#### Query Match

1.5%; Score 8; DB 2; Length 647;

Best Local Similarity 100.0%; Pred. No. 19;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 QALEEEEE 55

|||||||

Db 454 QALEEEEE 461

#### RESULT 15

AH1535

pyruvate-flavodoxin oxidoreductase homolog nifJ [imported] - *Listeria innocua* (strain C)

C;Species: *Listeria innocua*

C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001

C;Accession: AH1535

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker

; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl, H.

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A;Authors: Kraft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma

ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,

A;Title: Comparative genomics of *Listeria* species

A;Reference number: AB1077; MUID:21537279; PMID:11679669

A;Accession: AH1535

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1216 <GLA>

A;Cross-references: GB:AL592022; PIDN:CAC96056.1; PID:g16413275; GSPDB:GN00178

A;Experimental source: strain Clp11262

C;Genetics:

A;Gene: nifJ

C;Superfamily: pyruvate (flavodoxin) dehydrogenase; ferredoxin 2[4Fe-4S] homology

#### Query Match

1.5%; Score 8; DB 2; Length 1216;

Best Local Similarity 100.0%; Pred. No. 34;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 405 ETISEGLR 412

|||||||

Db 902 ETISEGLR 909

Search completed: July 23, 2002, 19:09:17

Job time: 303 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 19:07:55 ; Search time 16.34 Seconds  
(without alignments)  
1303.290 Million cell updates/sec

Title: US-09-495-823-7  
Perfect score: 550  
Sequence: 1 MAPRCAGHPPPSPQACV.....VPRYPKDRSPRLNGGV 550

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	2.0	473	1 ARSB_RAT	P50430 rattus norv
2	10	1.8	285	1 ARSB_MOUSE	P50429 mus musculu
3	10	1.8	533	1 ARSB_HUMAN	P15848 homo sapien
4	10	1.8	535	1 ARSB_FELCA	P33727 felis silve
5	8	1.5	176	1 PUR6_SYNY3	Q55498 synchocyst
6	8	1.5	551	1 ARS_HEMPU	P14000 hemiceintrot
7	8	1.5	567	1 ARS_STRPU	P50473 strongyloce
8	8	1.5	647	1 NOG1_YEAST	Q02892 saccharomyc
9	8	1.5	1420	1 SRB9_YEAST	P38931 saccharomyc
10	7	1.3	237	1 CBIL_SALTY	Q05593 salmonella
11	7	1.3	238	1 VGL6_HSV11	P06484 herpes simp
12	7	1.3	252	1 PSA3_ACACA	P90513 acanthamoeb
13	7	1.3	278	1 PANC_NEIMB	P57036 neisseria m
14	7	1.3	302	1 HSLO_SYNY3	P73910 synchocyst
15	7	1.3	304	1 TRUE_BACHD	Q9kab0 bacillus ha
16	7	1.3	339	1 XAP5_HUMAN	Q14320 homo sapien
17	7	1.3	387	1 RECF_CAUCR	P49998 caulobacter
18	7	1.3	391	1 MUCB_HUMAN	P04220 homo sapien
19	7	1.3	397	1 MML6_MYCTU	Q10773 mycobacteri
20	7	1.3	449	1 COMB_STRPN	P36498 streptococc
21	7	1.3	454	1 MUC_HUMAN	P01871 homo sapien
22	7	1.3	458	1 C6ST_CHICK	Q92179 gallus gall
23	7	1.3	464	1 ARS_KLEAE	P20713 klebsiella
24	7	1.3	470	1 EGR2_MOUSE	P08152 mus musculu
25	7	1.3	472	1 C901_ARATH	Q42569 arabidopsis
26	7	1.3	489	1 YHIP_ECOLI	P36837 escherichia
27	7	1.3	491	1 VIEL_HCMVA	P13202 human cytom
28	7	1.3	491	1 VIEL_HCMVT	P03169 human cytom
29	7	1.3	505	1 AINX_RAT	P23565 rattus norv
30	7	1.3	508	1 CPT7_CHICK	P12394 gallus gall
31	7	1.3	522	1 GA6S_HUMAN	P34059 homo sapien
32	7	1.3	537	1 YRR2_EBV	P03210 Epstein-Bar
33	7	1.3	550	1 LIGA_MOUSE	Q61211 mus musculu

34	7	1.3	551	1 YABN_ECOLI	P33595 escherichia
35	7	1.3	555	1 NFL_COTJA	Q02916 coturnix co
36	7	1.3	563	1 Y288_CHLTR	O84290 chlamydia t
37	7	1.3	577	1 STS_RAT	P15589 rattus norv
38	7	1.3	581	1 ILV5_PEA	O82043 pisum sativ
39	7	1.3	586	1 TUI2_SCHPO	Q9uuu8 schizosacch
40	7	1.3	597	1 GAN_HUMAN	O9h2c0 homo sapien
41	7	1.3	621	1 FOR_THELI	Q56303 thermococcu
42	7	1.3	641	1 HS7A_DROME	P29843 drosophila
43	7	1.3	646	1 NA95_HUMAN	Q9ulx6 homo sapien
44	7	1.3	652	1 NU5C_MESVI	Q9muk8 mesostigma
45	7	1.3	655	1 ILF1_HUMAN	Q01167 homo sapien

ALIGNMENTS

RESULT	1	ARSB_RAT	ID	ARSB_RAT	STANDARD;	PRT;	473 AA.
AC	P50430;						
DT	01-OCT-1996 (Rel. 34, Created)						
DT	01-OCT-1996 (Rel. 34, Last sequence update)						
DT	15-JUL-1998 (Rel. 36, Last annotation update)						
DE	Arylsulfatase B (EC 3.1.6.12) (ASB) (N-acetylgalactosamine-4-sulfatase) (G4S) (Fragment).						
DE	4-sulfatase) (G4S) (Fragment).						
GN	ARSB.						
OS	Rattus norvegicus (Rat).						
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.						
OX	NCBI_TaxID=101116;						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RC	STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;						
RX	MEDLINE=96121368; PubMed=8575749;						
RA	Kunieda T.;						
RT	*Mucopolysaccharidosis type VI in rats: isolation of cDNAs encoding arylsulfatase B, chromosomal localization of the gene, and identification of the mutation.;						
RL	Genomics 29:582-587(1995).						
CC	-I- CATALYTIC ACTIVITY: Hydrolysis of the 4-sulfate groups of the N-acetyl-D-galactosamine 4-sulfate units of chondroitin sulfate and dermatan sulfate.						
CC	-I- SUBUNIT: HOMODIMER (BY SIMILARITY).						
CC	-I- SUBCELLULAR LOCATION: Lysosomal.						
CC	-I- DISEASE: DEFECTS IN ARSB ARE THE CAUSE OF MARTEAUX-LAMY SYNDROME; ALSO KNOWN AS MUCOPOLYSACCHARIDOSIS TYPE VI (MPS-VI).						
CC	-I- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.						
CC	-----						
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).						
CC	-----						
DR	EMBL; D49434; BAA08412.1; -						
DR	HSSP; P15848; LFSU						
DR	InterPro; IPR000917; Sulfatase.						
DR	Pfam; PF00884; Sulfatase; 1.						
DR	PROSITE; PS00523; SULFATASE_1; 1.						
DR	PROSITE; PS00149; SULFATASE_2; FALSE_NEG.						
KW	Hydrolase; Glycoprotein; Lysosome; Mucopolysaccharidosis.						
FT	NON_TER	1	1				
FT	MOD_RES	31	31				
FT							
FT	ACT_SITE	87	87				
FT	POTENTIAL.						
FT	DISULFID	57	461				
FT	BY SIMILARITY.						
FT	DISULFID	61	95				
FT	BY SIMILARITY.						
FT	DISULFID	121	132				
FT	BY SIMILARITY.						
FT	DISULFID	345	387				
FT	BY SIMILARITY.						
FT	CARBOHYD	128	128				
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).						

FT CARBOHYD 219 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 231 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 355 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 366 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 398 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 473 AA; 53320 MW; 4E114C923A24AF8F CRC64;

Query Match 2.0%; Score 11; DB 1; Length 473;  
 Best Local Similarity 100.0%; Pred. No. 0.012;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 STMVGVKWHLG 180  
 |||||  
 DB 79 STMVGVKWHLG 89

## RESULT 2

ARSB\_MOUSE  
 ID ARSB\_MOUSE STANDARD; PRT; 285 AA.  
 AC P50429;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Arylsulfatase B (EC 3.1.6.12) (ASB) (N-acetylgalactosamine-  
 4-sulfatase) (G4S) (Fragments).  
 DE ARSB OR AS1.  
 GN Mus musculus (Mouse).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE OF 1-33 FROM N.A.  
 RX MEDLINE=92241876; PubMed=1572648;  
 RA Grompe M., Pieretti M., Caskey C.T., Ballabio A.;  
 RT "The sulfate gene family: cross-species PCR cloning using the MOPAC  
 RT technique.";  
 RL Genomics 12:755-760(1992).  
 RN [2]  
 RP SEQUENCE OF 34-285 FROM N.A.  
 RX MEDLINE=96323200; PubMed=8710849;  
 RA Evers M., Saftig P., Schmidt P., Hafner A., McLothlin D.B.,  
 RA Schmahl W., Hess B., von Figura K., Peters C.W.B.;  
 RT "Targeted disruption of the arylsulfatase B gene results in mice  
 RT resembling the phenotype of mucopolysaccharidosis VI.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:8214-8219(1996).  
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 4-sulfate groups of the N-  
 CC acetyl-D-galactosamine 4-sulfate units of chondroitin sulfate and  
 CC dermatan sulfate.  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Lysosomal.  
 CC -1- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M82877; AAA37261.1; .  
 CC EMBL; X92096; CAA63067.1; .  
 CC HSP; P15848; IFSU  
 CC MGD; MGI:88075; As1-s.  
 CC InterPro: IPR000917; Sulfatase.  
 CC Pfam; PF00884; Sulfatase; 1.  
 CC PROSITE; PS00523; SULFATASE\_1; PARTIAL.  
 CC PROSITE; PS00149; SULFATASE\_2; 1.  
 KW Hydrolase; Glycoprotein; Lysosome.  
 FT NON\_TER 1 1  
 FT NON\_CONS 33 34  
 FT ACT\_SITE 45 45 POTENTIAL.

FT CARBOHYD 86 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 177 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 189 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 264 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT NON\_TER 285  
 SQ SEQUENCE 285 AA; 31727 MW; 36BA5B98EA0770C9 CRC64;

Query Match 1.8%; Score 10; DB 1; Length 285;  
 Best Local Similarity 100.0%; Pred. No. 0.073;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 THMVGKWHLG 180  
 |||||  
 DB 38 THMVGKWHLG 47

## RESULT 3

ARSB\_HUMAN  
 ID ARSB\_HUMAN STANDARD; PRT; 533 AA.  
 AC P15848;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Arylsulfatase B precursor (EC 3.1.6.12) (ASB) (N-acetylgalactosamine-  
 4-sulfatase) (G4S).  
 DE ARSB.  
 GN Homo sapiens (Human).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=90153994; PubMed=2303452;  
 RA Peters C., Schmidt B., Rommerskirch W., Rupp K., Zuehlendorf M.,  
 RA Vangron M., Meyer H.E., Pohlmann R., von Figura K.;  
 RT "Phylogenetic conservation of arylsulfatases. cDNA cloning and  
 RT expression of human arylsulfatase B.";  
 RL J. Biol. Chem. 265:3374-3381(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90152677; PubMed=1968043;  
 RA Schuchman E.H., Jackson C.E., Desnick R.J.;  
 RT "Human arylsulfatase B: MOPAC cloning, nucleotide sequence of a full-  
 RT length cDNA, and regions of amino acid identity with arylsulfatases A  
 RT and C.";  
 RL Genomics 6:149-158(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93332648; PubMed=7687847;  
 RA Modaresi S., Rupp K., von Figura K., Peters C.;  
 RT "Structure of the human arylsulfatase B gene.";  
 RL Biol. Chem. Hoppe-Seyler 374:327-335(1993).  
 RN [4]  
 RP SEQUENCE OF 1-104 FROM N.A.  
 RX MEDLINE=92028992; PubMed=1930244;  
 RA Litjens T., Morris C.P., Gibson G.J., Beckmann K.R., Hopwood J.J.;  
 RT "Human N-acetylgalactosamine-4-sulphatase: protein maturation and  
 RT isolation of genomic clones.";  
 RL Biochem. Int. 24:209-215(1991).  
 RN [5]  
 RP 2-AMINO-3-OXOPROPIONIC ACID MODIFICATION SITE.  
 RX MEDLINE=95354208; PubMed=7628016;  
 RA Schmidt B., Selmer T., Ingendoh A., von Figura K.;  
 RT "A novel amino acid modification in sulfatases that is defective in  
 RT multiple sulfatase deficiency.";  
 RL Cell 82:271-278(1995).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
 RX MEDLINE=97184692; PubMed=9032078;  
 RA Bond C.S., Clements P.R., Ashby S.J., Collyer C.A., Harrop S.J.,  
 RA Hopwood J.J., Guss J.M.;  
 RT "Structure of a human lysosomal sulfatase.";

RL Structure 5:277-289(1997).  
RN [7]  
RP VARIANTS MPS-VI ARG-117; PRO-236 AND TYR-405.  
RX MEDLINE-92197625; PubMed-1550123;  
RA Jin W.-D., Jackson C.E., Desnick R.J., Schuchman E.H.;  
RT "Mucopolysaccharidosis type VI: identification of three mutations in  
the arylsulfatase B gene of patients with the severe and mild  
phenotypes provides molecular evidence for genetic heterogeneity.";  
RL Am. J. Hum. Genet. 50:795-800(1992).  
RN [8]  
RP VARIANT MPS-VI VAL-137, AND VARIANT MET-376.  
RX MEDLINE-92042029; PubMed-1718978;  
RA Wicker G., Prill V., Brooks D., Gibson G., Hopwood J.,  
von Figura K., Peters C.;  
RT "Mucopolysaccharidosis VI (Maroteaux-Lamy syndrome). An intermediate  
clinical phenotype caused by substitution of valine for glycine at  
position 137 of arylsulfatase B.";  
RL J. Biol. Chem. 266:21386-21391(1991).  
RN [9]  
RP VARIANTS MPS-VI MET-92; GLN-95; CYS-210; PRO-393 AND PRO-498.  
RX MEDLINE-96213747; PubMed-8651289;  
RA Litjens T., Brooks D.A., Peters C., Gibson G.J., Hopwood J.J.;  
RT "Identification, expression, and biochemical characterization of N-  
acetylgalactosamine-4-sulfatase mutations and relationship with  
clinical phenotype in MPS-VI patients.";  
RL Am. J. Hum. Genet. 58:1127-1134(1996).  
RN [10]  
RP VARIANTS MPS-VI TRP-152 AND GLN-160.  
RX MEDLINE-94171224; PubMed-8125475;  
RA Voskoboeva E., Isbrandt D., von Figura K., Krasnopolskaya X.,  
Peters C.;  
RT "Four novel mutant alleles of the arylsulfatase B gene in two  
patients with intermediate form of mucopolysaccharidosis VI  
(Maroteaux-Lamy syndrome).";  
RL Hum. Genet. 93:259-264(1994).  
RN [11]  
RP VARIANT MPS-IV ARG-302.  
RA Villani G.R.D., Balzano N., di Natale P.;  
RT "Two novel mutations of the arylsulfatase B gene in two Italian  
patients with severe form of mucopolysaccharidosis.";  
RL Hum. Mutat. 11:410-410(1998).  
CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 4-sulfate groups of the N-  
acetyl-D-galactosamine 4-sulfate units of chondroitin sulfate and  
dermatan sulfate.  
CC -1- SUBUNIT: MONOMER.  
CC -1- SUBCELLULAR LOCATION: Lysosomal.  
CC -1- DISEASE: DEFECTS IN ARSB ARE THE CAUSE OF MAROTEAUX-LAMY  
SYNDROME; ALSO KNOWN AS MUCOPOLYSACCHARIDOSIS TYPE VI (MPS-VI).  
CC THIS DISEASE IS CHARACTERIZED BY THE ACCUMULATION OF DERMATAN  
SULFATE IN LYSSOMES. CLINICAL FEATURES CAN INCLUDE ABNORMAL  
GROWTH, SHORT STATURE, STIFF JOINTS, SKELETAL MALFORMATIONS,  
CORNEAL CLOUDING, HEPATOSPLENOMEGALY, AND CARDIAC ABNORMALITIES.  
CC A WIDE VARIATION IN CLINICAL SEVERITY IS OBSERVED.  
CC -1- DISEASE: MULTIPLE SULFATASE DEFICIENCY (MSD) IS A DISORDER THAT  
COMBINES FEATURES OF METACHROMATIC LEUKODYSTROPHY AND OF  
MUCOPOLYSACCHARIDOSIS; IT IS CHARACTERIZED BY A DECREASED  
ACTIVITY OF ALL KNOWN SULFATASES. IT SEEMS TO BE CAUSED FROM THE  
LACK OF POST-TRANSLATIONAL MODIFICATION OF A CYSTEINE INTO 2-  
AMINO-3-OXOPROPIONIC ACID.  
CC -1- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; J05225; AAA51784.1; -;  
DR EMBL; M32373; AAA51779.1; -;  
DR EMBL; X72735; CAA51272.1; -;  
DR EMBL; X72736; CAA51272.1; JOINED.

DR EMBL; X72737; CAA51272.1; JOINED.  
DR EMBL; X72738; CAA51272.1; JOINED.  
DR EMBL; X72739; CAA51272.1; JOINED.  
DR EMBL; X72740; CAA51272.1; JOINED.  
DR EMBL; X72741; CAA51272.1; JOINED.  
DR EMBL; X72742; CAA51272.1; JOINED.  
DR EMBL; S57777; AAB19988.1; -;  
DR PIR; A35078; A35078.  
DR PDB; 1FSU; 04-FEB-98.  
DR MIM; 253200; -;  
DR MIM; 272200; -;  
DR InterPro; IPR000917; Sulfatase.  
DR Pfam; PF00884; Sulfatase; 1.  
DR PROSITE; PS00523; SULFATASE\_1; 1.  
DR PROSITE; PS00149; SULFATASE\_2; 1.  
KW Hydrolase; Signal; Glycoprotein; Lysosome; Mucopolysaccharidosis;  
KW Disease mutation; Polymorphism; 3D-structure;  
FT SIGNAL 1 36  
OR 38 (POTENTIAL).  
FT CHAIN 37 533  
ARYLSULFATASE B.  
FT ACT\_SITE 147 147  
POTENTIAL.  
FT MOD\_RES 91 91  
2-AMINO-3-OXOPROPIONIC ACID.  
FT DISULFID 117 521  
FT DISULFID 121 155  
FT DISULFID 181 192  
FT DISULFID 405 447  
FT CARBOHYD 188 188  
FT CARBOHYD 279 279  
FT CARBOHYD 291 291  
FT CARBOHYD 366 366  
FT CARBOHYD 426 426  
FT CARBOHYD 458 458  
FT VARIANT 92 92  
N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT VARIANT 95 95  
R -> Q (IN MPS-VI; MILD/SEVERE FORM).  
FT VARIANT 117 117  
/FTID-VAR\_007295.  
FT VARIANT 137 137  
G -> V (IN MPS-VI; SEVERE FORM).  
FT VARIANT 152 152  
R -> W (IN MPS-VI; INTERMEDIATE FORM).  
FT VARIANT 160 160  
R -> Q (IN MPS-VI; INTERMEDIATE FORM).  
FT VARIANT 210 210  
Y -> C (IN MPS-VI; MILD/INTERMEDIATE).  
FT VARIANT 236 236  
L -> P (IN MPS-VI; MILD FORM).  
FT VARIANT 302 302  
G -> R (IN MPS-VI; SEVERE FORM).  
FT VARIANT 376 376  
V -> M.  
FT VARIANT 393 393  
H -> P (IN MPS-VI; MILD/SEVERE FORM).  
FT VARIANT 405 405  
C -> Y (IN MPS-VI; MILD FORM).  
FT VARIANT 498 498  
L -> P (IN MPS-VI; MILD/SEVERE FORM).  
FT CONFLICT 358 358  
V -> M (IN REF. 3).  
FT SEQUENCE 533 AA; 59687 MW; 5983FB691C4789A CRC64;

Query Match 1.8%; Score 10; DB 1; Length 533;  
Best Local Similarity 100.0%; Pred. No. 0.13;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 THMVGKWHLG 180  
|||||  
Db 140 THMVGKWHLG 149

RESULT 4  
ARSB\_FELCA STANDARD; PRY; 535 AA.  
ID ARSB\_FELCA

```

AC P3327;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Arylsulfatase B precursor (EC 3.1.6.12) (ASB) (N-acetylgalactosamine-
DE 4-sulfatase) (G4S).
GN ARSB.
OS Felis silvestris catus (Cat).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93052342; PubMed=1427856;
RA Jackson C.E., Yuhki N., Desnick R.J., Haskins M.E., O'Brien S.J.,
RA Schuchman E.H.;
RT "Feline arylsulfatase B (ARSB): isolation and expression of the cDNA,
RT comparison with human ARSB, and gene localization to feline
RT chromosome A1.";
RL Genomics 14:403-411(1992).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of the 4-sulfate groups of the N-
CC acetyl-D-galactosamine 4-sulfate units of chondroitin sulfate and
CC dermatan sulfate.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Lysosomal.
CC -1- DISUBS: DEFECTS IN ARSB ARE THE CAUSE OF MAROTEAUX-LAMY
CC SYNDROME; ALSO KNOWN AS MUCOPOLYSACCHARIDOSIS TYPE VI (MPS-VI).
CC MPS-VI HAS BEEN DESCRIBED IN SIAMESE CATS.
CC -1- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL: S48472; AAB23941.1; -.
DR PIR: A44475; A44475.
DR HSSP: P15848; IfSU.
DR InterPro: IPR000917; Sulfatase.
DR Pfam: PF00884; Sulfatase; 1.
DR PROSITE: PS00523; SULFATASE_1; 1.
DR PROSITE: PS00149; SULFATASE_2; 1.
KW Hydrolase; Signal; Glycoprotein; Lysosome; Mucopolysaccharidosis.
FT SIGNAL 1 41
FT CHAIN 42 535
FT ACT_SITE 149 149
FT MOD_RES 93 93
FT FT
FT DISULFID 119 523
FT DISULFID 123 157
FT DISULFID 183 194
FT DISULFID 407 449
FT CARBOHYD 190 190
FT CARBOHYD 281 281
FT CARBOHYD 293 293
FT CARBOHYD 428 428
FT CARBOHYD 460 460
FT SEQUENCE 535 AA; 59753 MW; 43A527886A9983C4 CRC64;
Query Match 1.8%; Score 10; DB 1; Length 535;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 171 THWVGKWHLG 180
DB 142 THWVGKWHLG 151

```

```

RESULT 5
PUR6_SYNY3
ID PUR6_SYNY3 STANDARD; PRT; 176 AA.
AC Q55498;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phosphoribosylaminoimidazole carboxylase catalytic subunit
DE (EC 4.1.1.21) (AIR carboxylase) (AIRC).
GN PURE OR SL0901.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
CC -1- FUNCTION: THIS SUBUNIT CAN ALONE TRANSFORM AIR TO CAIR, BUT IN
CC ASSOCIATION WITH PURK, WHICH POSSESSES AN ATPASE ACTIVITY, AN
CC ENZYME COMPLEX IS PRODUCED WHICH IS CAPABLE OF CONVERTING AIR TO
CC CAIR EFFICIENTLY UNDER PHYSIOLOGICAL CONDITION (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 1-(5-phosphoribosyl)-5-amino-4-
CC imidazolecarboxylate - 1-(5-phosphoribosyl)-5-aminoimidazole +
CC CO(2).
CC -1- PATHWAY: SIXTH STEP IN DE NOVO PURINE BIOSYNTHESIS.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SIMILARITY: TO OTHER BACTERIAL PURE, ALSO TO AIR CARBOXYLASE FROM
CC FUNGI.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL: D64006; BAA10848.1; -.
DR HSSP: P09028; IQC2.
DR InterPro: IPR000031; AIR_carboxyl.
DR Pfam: PF00731; AIRC; 1.
KW Purine biosynthesis; Lyase; Decarboxylase; Complete proteome.
SQ SEQUENCE 176 AA; 18603 MW; A1F87A630F18F8E4 CRC64;
Query Match 1.5%; Score 8; DB 1; Length 176;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 248 QILASHNP 255
DB 131 QILASHNP 138
RESULT 6
ARS_HEMPU
ID ARS_HEMPU STANDARD; PRT; 551 AA.
AC P14000;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Arylsulfatase precursor (EC 3.1.6.1) (Aryl-sulfate sulphohydrolase)
DE (ARS).
OS Hemientrotus pulcherrimus (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Hemientrotus.
OX NCBI_TaxID=7650;

```

RN RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Pluteus;  
 RX MEDLINE=89030699; PubMed=3181160;  
 RA Sasaki H., Yamada K., Akasaka H., Suzuki K., Saito A., Sato M.,  
 RA Shimada H.;  
 RT "cDNA cloning, nucleotide sequence and expression of the gene for  
 RT arylsulfatase in the sea urchin (*Hemicentrotus pulcherrimus*)  
 RL embryo.";  
 RL Eur. J. Biochem. 177:9-13(1988).  
 [2]  
 RN RP SEQUENCE FROM N.A.  
 RX MEDLINE=90092130; PubMed=2598936;  
 RA Yamada K., Akasaka K., Shimada H.;  
 RT "Structure of sea-urchin arylsulfatase gene.";  
 RL Eur. J. Biochem. 186:405-410(1989).  
 CC -1- FUNCTION: MAY BE A STRUCTURAL COMPONENT OF THE EXTRACELLULAR  
 CC MATRICES INVOLVED IN CELL MOVEMENT DURING MORPHOGENESIS.  
 CC -1- CATALYTIC ACTIVITY: A phenol sulfate + H(2)O = a phenol + sulfate.  
 CC -1- SUBCELLULAR LOCATION: IN BOTH THE CYTOPLASM AND THE EXTRACELLULAR  
 CC MATRICES.  
 CC -1- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X17015; CAA34881.1; -.  
 DR PIR; S01793; S01793.  
 DR PIR; S07089; S07089.  
 DR HSSP; P15289; IAUK.  
 DR InterPro; IPR000917; Sulfatase.  
 DR Pfam; PF00884; Sulfatase; 1.  
 DR PROSITE; PS00523; SULFATASE\_1; 1.  
 DR PROSITE; PS00149; SULFATASE\_2; 1.  
 KW Hydrolase; Signal; Glycoprotein; Extracellular matrix.  
 FT SIGNAL 1 20  
 FT CHAIN 21 551  
 FT MOD\_RES 21 21  
 FT MOD\_RES 100 100  
 FT 2-AMINO-3-OXOPROPIONIC ACID (BY  
 FT BLOCKED.  
 FT ACT\_SITE 158 158  
 FT CARBOHYD 164 164  
 FT CARBOHYD 213 213  
 FT CARBOHYD 296 296  
 FT SEQUENCE 551 AA; 60952 MW; 54C1AAC14D6710C9 CRC64;  
 SQ  
 Query Match 1.5%; Score 8; DB 1; Length 551;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 173 MVGKWHLG 180  
 Db 153 MVGKWHLG 160  
 RESULT 7  
 ARS\_STRPU  
 ID ARS\_STRPU STANDARD; PRT; 567 AA.  
 AC P50473;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Arylsulfatase precursor (EC 3.1.6.1) (Aryl-sulfate sulphohydrolase)  
 DE (ARS).  
 OS Strongylocentrotus purpuratus (Purple sea urchin).  
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 OC Echinoidea; Echinoidea; Echinacea; Echinoida; Strongylocentrotidae;

OC Strongylocentrotus.  
 OX NCBI\_TaxID=7668;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89357267; PubMed=2767335;  
 RA Yang Q., Angerer L.M., Angerer R.C.;  
 RA "Structure and tissue-specific developmental expression of a sea  
 RT urchin arylsulfatase gene.";  
 RL Dev. Biol. 135:53-65(1989).  
 CC -1- CATALYTIC ACTIVITY: A phenol sulfate + H(2)O = a phenol + sulfate.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: EXPRESSION IS CONFINED TO ABORAL ECTODERM  
 CC CELLS AND THEIR PRECURSORS.  
 CC -1- DEVELOPMENTAL STAGE: LOW LEVELS ARE FOUND AT MESENCHYME BLASTULA  
 CC STAGE (24 HR), LEVELS INCREASE BY LATE GASTRULA STAGE AND ARE  
 CC MAINTAINED AT PLUTEUS STAGE.  
 CC -1- SIMILARITY: BELONGS TO THE SULFATASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M28404; AAA30036.1; -.  
 DR HSSP; P15289; IAUK.  
 DR InterPro; IPR000917; Sulfatase.  
 DR Pfam; PF00884; Sulfatase; 1.  
 DR PROSITE; PS00523; SULFATASE\_1; 1.  
 DR PROSITE; PS00149; SULFATASE\_2; 1.  
 KW Hydrolase; Signal; Glycoprotein.  
 FT SIGNAL 1 19  
 FT CHAIN 20 567  
 FT MOD\_RES 115 115  
 FT 2-AMINO-3-OXOPROPIONIC ACID (BY  
 FT SIMILARITY).  
 FT ACT\_SITE 173 173  
 FT CARBOHYD 179 179  
 FT CARBOHYD 228 228  
 FT CARBOHYD 542 542  
 FT SEQUENCE 567 AA; 62477 MW; D55B627983A4C4D6 CRC64;  
 SQ  
 Query Match 1.5%; Score 8; DB 1; Length 567;  
 Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 173 MVGKWHLG 180  
 Db 168 MVGKWHLG 175  
 RESULT 8  
 NOGI\_YEAST  
 ID NOGI\_YEAST STANDARD; PRT; 647 AA.  
 AC Q02892;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Nucleolar GTP-binding protein 1.  
 GN NOGI OR YPL093W OR LPG15W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RX MEDLINE=97313271; PubMed=9169875;  
 RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoorge W.,  
 RA Araujo R., Aparicio A., Barrell B., Badcock K., Benes V., Botstein D.,  
 RA Bowman S., Bruckner M., Carpenter J., Cherry J.M., Chung E.,

RA Churcher C., Coster F., Davis K., Davis R.W., Dietrich F.S.,  
RA Delius H., DiPaolo T., Dubois E., Dusterhoft A., Duncan M., Floeth M.,  
RA Fortin N., Friesen J.D., Fritz C., Goffeau A., Hall J., Hebling U.,  
RA Heumann K., Hilbert H., Hillier L., Hunnicke-Smith S., Hyman R., D.,  
RA Johnston M., Kalman S., Kleinke K., Komp C., Kordl O., Lashkari D.,  
RA Lew H., Lin A., Lin D., Louis E.J., Marathe R., Messenguy F.,  
RA Mewes H.W., Mirtipati S., Moestl D., Muller-Auer S., Namath A.,  
RA Newtlich U., Oefner P., Pearson D., Petel F.X., Pohl T.M.,  
RA Purnelle D., Schafer M., Scharfe M., Scherens B., Schramm S.,  
RA Schroeder M., Seicu A.M., Tettelin H., Urrestarazu L.A., Ushinsky S.,  
RA Vierdeels F., Viessers S., Voss H., Walsh S.V., Wambutt R., Wang Y.,  
RA Wedler E., Wedler H., Winnett E., Zhong W.W., Zollner A., Vo D.H.,  
RA Hani J.,  
RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI";  
RA Nature 387:103-105(1997).  
RA [2]  
RA SUBCELLULAR LOCATION.  
RX MEDLINE-21062923; PubMed-11112701;  
RA Park J.-H., Jensen B.C., Kifer C.T., Parsons M.;  
RT "A novel nucleolar G-protein conserved in eukaryotes.";  
RL J. Cell Sci. 114:173-185(2001).  
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar.  
CC -1- SIMILARITY: BELONGS TO THE GTP1 / OBG FAMILY. NOG SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U43281; AAB68206.1; -;  
DR EMBL; U09176; AAA18614.1; -;  
DR EMBL; U33007; AAB64875.1; -;  
DR PIR; S48538; S48538.  
DR TRANSPAC; T02153; -;  
DR SGD; S0002851; SSN2.  
KW Nuclear protein.  
FT DOMAIN 526 529  
FT DOMAIN 526 664  
FT DOMAIN 657 664  
FT DOMAIN 813 816  
FT DOMAIN 1005 1008  
FT DOMAIN 1121 1136  
FT CONFLICT 38 38  
FT CONFLICT 812 812  
FT CONFLICT 859 859  
FT CONFLICT 877 878  
FT CONFLICT 887 887  
FT CONFLICT 1284 1284  
SQ SEQUENCE 647 AA; 74409 MW; 640324779AB4D716 CRC64;  
Query Match 1.5%; Score 8; DB 1; Length 647;  
Best Local Similarity 100.0%; Pred. No. 13;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 48 QALEEEEE 55  
DB 454 QALEEEEE 461  
|||||||  
RESULT 9  
SRB9\_YEAST  
ID SRB9\_YEAST STANDARD; PRT; 1420 AA.  
AC P38931;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DE 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Suppressor of RNA polymerase B SRB9 (SCA1 protein).  
GN SRB9 OR SCA1 OR SSN2 OR YDR443C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-S288C;  
RX MEDLINE-95293223; PubMed=7774808;  
RA Hengartner C.J., Thompson C.M., Zhang J., Chao D.M., Liao S.-M.,  
RA Koleske A.J., Okamura S., Young R.A.;  
RT "Association of an activator with an RNA polymerase II holoenzyme.";  
RL Genes Dev. 9:897-910(1995).  
RN [2]  
RP SEQUENCE FROM N.A.

RA Yuryev A., Corden J.L.;  
RA Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.  
RL [3]  
RP SEQUENCE FROM N.A.  
RA Dietrich F.S., Mulligan J., Allen E., Araujo R., Aviles E., Berno A.,  
RA Carpenter J., Chen E., Cherry J.M., Chung E., Duncan M.,  
RA Hunnicke-Smith S., Hyman R., Komp C., Lashkari D., Lew H., Lin D.,  
RA Mosedale D., Nakahara K., Namath A., Oefner P., Oh C., Petel F.X.,  
RA Roberts D., Schramm S., Schroeder M., Shogren T., Shroff N.,  
RA Winant A., Yelton M., Botstein D., Davis R.W.;  
RA Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
RL Submitter (AUG-1995) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: COMPONENT OF THE RNA POLYMERASE II HOLOENZYME AND THE  
CC MEDIATOR OF ACTIVATION SUBCOMPLEX.  
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U23812; AAA91316.1; -;  
DR EMBL; U09176; AAA18614.1; -;  
DR EMBL; U33007; AAB64875.1; -;  
DR PIR; S48538; S48538.  
DR TRANSPAC; T02153; -;  
DR SGD; S0002851; SSN2.  
KW Nuclear protein.  
FT DOMAIN 526 529  
FT DOMAIN 526 664  
FT DOMAIN 657 664  
FT DOMAIN 813 816  
FT DOMAIN 1005 1008  
FT DOMAIN 1121 1136  
FT CONFLICT 38 38  
FT CONFLICT 812 812  
FT CONFLICT 859 859  
FT CONFLICT 877 878  
FT CONFLICT 887 887  
FT CONFLICT 1284 1284  
SQ SEQUENCE 1420 AA; 160000 MW; 7F6CF4BBE0FAC918 CRC64;  
Query Match 1.5%; Score 8; DB 1; Length 1420;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 153 LPDNDATL 160  
DB 1033 LPDNDATL 1040  
|||||||  
RESULT 10  
CBIL\_SALTY  
ID CBIL\_SALTY STANDARD; PRT; 237 AA.  
AC Q05593;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DE 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Precorrin-2 C20-methyltransferase (EC 2.1.1.130) (S-adenosyl-L-  
DE methionine--precorrin-2 methyltransferase) (SP2MT).  
GN CBIL OR STM2024  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF N-TERMINUS.  
RC STRAIN-L72;  
RX MEDLINE-93273696; PubMed=8501034;  
RA Roth J.R., Lawrence J.G., Rubinfeld M., Kieffer-Higgins S.,  
RA Church G.M.;



```

RT *Characterization of the cobalamin (vitamin B12) biosynthetic genes
RT of Salmonella typhimurium.
RL J. Bacteriol. 175:3303-3316(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=J2 / SCS1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA *Complete genome sequence of Salmonella enterica serovar Typhimurium
RT J2;
RL Nature 413:852-856(2001).
RC -1- FUNCTION: METHYLATES PRECORRIN-2 AT THE C-20 POSITION TO PRODUCE
CC PRECORRIN-3A.
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + precorrin-2 -> S-
CC adenosyl-L-homocysteine + precorrin-3A.
CC -1- PATHWAY: COBALAMIN BIOSYNTHESIS.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS SUMT, CYSG, CBIF/COBM
CC AND CBIL/COBI.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L12006; AAA27263.1; -
CC EMBL; AE008789; AAL20928.1; -
CC StyGene; SG10044; cbil.
CC InterPro; IPR000878; Corrin_porph_methyltrnf.
CC InterPro; IPR003043; Uropor_methyltransf.
CC Pfam; PF00590; TP_methylase; 1.
CC PROSITE; PS00839; SUMT.1; 1.
CC PROSITE; PS00840; SUMT.2; 1.
CC Cobalamin biosynthesis; Porphyrin biosynthesis; Transferase;
CC Methyltransferase; Complete proteome.
CC CONFLICT 188 188 Q -> T (IN REF. 1).
CC SEQUENCE 237 AA; 25806 MW; 3D7215BDC67D2067 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 237;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 ALLAAQAG 63
DB 184 ALLAAQAG 190
|||||||

RESULT 11
VGLG_HSV11
ID VGLG_HSV11 STANDARD; PRT; 238 AA.
AC P06484;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Glycoprotein G.
GN GG OR US4.
OS Herpes simplex virus (type 1 / strain 17).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85160822; PubMed=2984429;
RA McGeoch D.J., Dolan A., Donald S., Rixon F.J.;
RA *Sequence determination and genetic content of the short unique
RT

```

```

RT region in the genome of herpes simplex virus type 1.
RL J. Mol. Biol. 181:1-13(1985).
CC -1- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1 AND
CC 2: GH, GB, GC, GG, GD, GI, AND GE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L00036; AAA96684.1; -
CC EMBL; X14112; CAA32281.1; -
CC EMBL; X02138; CAA26058.1; -
CC PIR; A05239; Q0BE74.
CC Glycoprotein. 28 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 49 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 49
CC SQ SEQUENCE 238 AA; 25238 MW; F7FDC2867E834B92 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 238;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 EEEEEGA 57
DB 81 EEEEEGA 87
|||||||

RESULT 12
PSA3_ACACA
ID PSA3_ACACA STANDARD; PRT; 252 AA.
AC P90513;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Proteasome subunit alpha type 3 (EC 3.4.25.1) (Fragment).
OS Acanthamoeba castellanii (Amoeba).
OC Eukaryota; Acanthamoebidae; Acanthamoeba.
OX NCBI_TaxID=5755;
RN [1]
RP SEQUENCE FROM N.A.
RA Xu P., Zot H.G.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE PROTEASOME IS A MULTICATALYTIC PROTEINASE COMPLEX
CC WHICH IS CHARACTERIZED BY ITS ABILITY TO CLEAVE PEPTIDES WITH ARG,
CC PHE, TYR, LEU, AND GLU ADJACENT TO THE LEAVING GROUP AT NEUTRAL OR
CC SLIGHTLY BASIC PH. THE PROTEASOME HAS AN ATP-DEPENDENT PROTEOLYTIC
CC ACTIVITY.
CC -1- CATALYTIC ACTIVITY: Cleavage at peptide bonds with very broad
CC specificity.
CC -1- PATHWAY: Involved in an ATP/ubiquitin-dependent non-lysosomal
CC proteolytic pathway.
CC -1- SUBUNIT: THE PROTEASOME IS COMPOSED OF AT LEAST 15 NON IDENTICAL
CC SUBUNITS WHICH FORM A HIGHLY ORDERED RING-SHAPED STRUCTURE (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic and nuclear (By similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1A; ALSO KNOWN AS THE
CC PROTEASOME A-TYPE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U85398; AAB41645.1; -
CC HSSP; P25156; IPMA.

```

```

DR  MEROPS; T01.977; .
DR  InterPro; IPR001353; Proteasome.
DR  InterPro; IPR000426; Proteasome_A.
DR  Pfam; PF00227; proteasome; 1.
DR  PROSITE; PS00388; PROTEASOME_A; 1.
KW  Proteasome; Hydrolase; Protease.
FT  NON_TER 1
SQ  SEQUENCE 252 AA; 27728 MW; 71898276850C2E79 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 252;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 QALEEE 54
    |||||
DB 239 QALEEE 245

RESULT 13
PANC_NEIMB
ID PANC_NEIMB STANDARD; PRT; 278 AA.
AC P57036;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Pantoate-beta-alanine ligase (EC 6.3.2.1) (pantothenate synthetase)
DE (Pantoate activating enzyme).
GN PANC OR NMB0871.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=2017575; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Citron H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58";
RL Science 287:1809-1815(2000).
CC -1- CATALYTIC ACTIVITY: ATP + (R)-pantoate + beta-alanine = AMP +
CC diphosphate + (R)-pantothenate.
CC -1- PATHWAY: Pantothenate biosynthesis; last step.
CC -1- SIMILARITY: BELONGS TO THE PANTOTHENATE SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE002439; AAP41282.1; .
DR TIGR; NMB0871; .
DR InterPro; IPR003721; Pantoate_ligase.
DR Pfam; PF02569; Pantoate_ligase; 1.
KW Pantothenate biosynthesis; Ligase; Complete proteome.
SQ SEQUENCE 278 AA; 31080 MW; 5E71D92FE0E6F6B3 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 104 DKLAEG 110

```

```

DB 76 DKLAEG 82
    |||||
RESULT 14
HSLO_SYNY3
ID HSLO_SYNY3 STANDARD; PRT; 302 AA.
AC P73910;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 33 kDa chaperonin (Heat shock protein 33 homolog) (HSP33).
GN HSLO OR SLL1988.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hikosawa M., Suglura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- FUNCTION: REDOX REGULATED MOLECULAR CHAPERONE. PROTECTS BOTH
CC THERMALLY UNFOLDING AND OXIDATIVELY DAMAGED PROTEINS FROM
CC IRREVERSIBLE AGGREGATION. PLAYS AN IMPORTANT ROLE IN THE BACTERIAL
CC DEFENSE SYSTEM TOWARD OXIDATIVE STRESS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- PTM: UNDER OXIDIZING CONDITIONS TWO DISULFIDE BONDS ARE FORMED
CC INVOLVING THE REACTIVE CYSTEINES. UNDER REDUCING CONDITIONS ZINC
CC IS BOUND TO THE REACTIVE CYSTEINES AND THE PROTEIN IS INACTIVE (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HSP33 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90910; BAA1974.1; .
DR InterPro; IPR000397; Hsp33.
DR Pfam; PF01430; HSP33; 1.
DR ProDom; PD014025; Hsp33; 1.
KW Chaperone; redox-active center; Zinc; Complete proteome.
FT DISULFID 240 242 REDOX-ACTIVE (BY SIMILARITY).
FT DISULFID 273 276 REDOX-ACTIVE (BY SIMILARITY).
SQ SEQUENCE 302 AA; 32205 MW; 0EB3448B06E7AAA CRC64;

Query Match 1.3%; Score 7; DB 1; Length 302;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 351 GGIRAVG 357
    |||||
DB 13 GGIRAVG 19

RESULT 15
TRUB_BACHD
ID TRUB_BACHD STANDARD; PRT; 304 AA.
AC Q9KA80;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

DE tRNA pseudouridine synthase B (EC 4.2.1.70) (tRNA pseudouridine 55  
DE synthase) (Pse55 synthase) (pseudouridylylate synthase) (Uracil  
DE hydrolase).  
GN TRUB OR BH2410.  
OS Bacillus halodurans.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-125/JCM 9153;  
RX MEDLINE=20512582; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,  
RA Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
RT halodurans and genomic sequence comparison with Bacillus subtilis.";  
RL Nucleic Acids Res. 28:4317-4331(2000).  
CC -!- FUNCTION: FORMATION OF PSEUDOURIDINE AT POSITION 55 IN THE  
CC PSI GC LOOP OF TRANSFER RNAs (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine  
CC 5'-phosphate + H(2)O.  
CC -!- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF001515; BAB06129.1; -.  
DR InterPro: IPR002501; TruB\_N.  
DR Pfam: PF01509; TruB\_N; 1.  
KW Lyase; tRNA processing; Complete proteome.  
SQ SEQUENCE 304 AA; 34054 MW; 2DE397CE20C67946 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 304;  
Best Local Similarity 100.0%; Pred. No. 61;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 EGGALLA 60  
|  
Db 275 EGGALLA 281

Search completed: July 23, 2002, 19:10:27  
Job time: 152 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 19:06:45 ; Search time 28.8 seconds  
(without alignments)  
3303.723 Million cell updates/sec

Title: US-09-495-823-7  
Perfect score: 550  
Sequence: 1 MAPRCAGHPPPSPQACVC.....VPRYPKPRSNRLNGGV 550

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63	11.5	169	4 Q9H5D8	Q9h5d8 homo sapien
2	20	3.6	266	5 Q9NJU7	Q9nJU7 helix pomat
3	12	2.2	503	5 Q9NJU8	Q9nJU8 helix pomat
4	8	1.5	217	5 Q9VQ95	Q9vQ95 drosophila
5	8	1.5	330	16 Q9HVR2	Q9hVR2 pseudomonas
6	8	1.5	381	16 Q92EX3	Q92ex3 listeria in
7	8	1.5	452	5 Q18924	Q18924 caenorhabdi
8	8	1.5	433	16 Q9CKE0	Q9cke0 pasteurella
9	8	1.5	542	5 Q9V6F2	Q9v6f2 drosophila
10	8	1.5	551	5 Q25384	Q25384 hemicirot
11	8	1.5	559	5 Q16138	Q16138 heliocidari
12	8	1.5	579	5 Q9VVM4	Q9vVM4 drosophila
13	8	1.5	898	2 Q93D92	Q93d92 streptococ
14	8	1.5	936	5 Q9VVK9	Q9vVK9 drosophila
15	8	1.5	1002	10 Q9AXB1	Q9axb1 oryza sativ
16	8	1.5	1216	16 Q92DI9	Q92di9 listeria in

17 1.3 24 4 Q96I53  
18 1.3 27 15 O11847  
19 1.3 48 5 Q9VWS6  
20 1.3 79 12 Q68100  
21 1.3 105 8 Q9B267  
22 1.3 114 16 O66649  
23 1.3 123 16 Q9K8T2  
24 1.3 123 17 Q9Y992  
25 1.3 126 12 O65435  
26 1.3 128 8 O78764  
27 1.3 128 12 Q68084  
28 1.3 141 16 Q98CC0  
29 1.3 143 4 Q96K66  
30 1.3 158 12 Q91TP8  
31 1.3 159 16 Q92DS6  
32 1.3 163 5 Q9NB86  
33 1.3 164 2 P71285  
34 1.3 169 2 O50216  
35 1.3 169 2 O33496  
36 1.3 169 2 O06836  
37 1.3 172 12 Q91D82  
38 1.3 172 15 Q9ORM7  
39 1.3 175 10 Q9FM63  
40 1.3 179 10 Q94LB7  
41 1.3 180 2 Q01092  
42 1.3 194 5 Q9U933  
43 1.3 198 3 P87027  
44 1.3 199 3 O14312  
45 1.3 199 4 Q9BUI1

Q96I53 homo sapien  
O11847 human immun  
Q9VWS6 drosophila  
Q68100 human cytom  
Q9B267 pachydaetyl  
O66649 aquifex aeo  
Q9K8T2 bacillus ha  
Q9Y992 aeropyrum p  
O65435 beet curly  
O78764 pyxis plani  
Q68084 human cytom  
Q98CC0 rhizobium l  
Q96K66 homo sapien  
Q91TP8 tupaia herp  
Q92DS6 listeria in  
Q9NB86 agrotis ips  
P71285 escherichia  
O50216 pseudomonas  
O33496 pseudomonas  
O06836 pseudomonas  
Q91D82 human cytom  
Q9ORM7 human immun  
Q9FM63 arabisopsis  
Q94LB7 oryza sativ  
Q01092 serratia ma  
Q9U933 paramecium  
P87027 schizosacch  
O14312 schizosacch  
Q9BUI1 homo sapien

#### ALIGNMENTS

RESULT 1

Q9H5D8 ID Q9H5D8 PRELIMINARY; PRT; 169 AA.  
AC Q9H5D8;  
DT 01-MAR-2001 (T-EMBLrel. 16, Created)  
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)  
DE CDNA: FLJ23548 FIS, CLONE LMG08487.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LUNG;  
RA Kawakami T., Nozuchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,  
RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,  
RA Yamada K., Fujii Y., Ozaki K., Hiraio M., Ohmori Y., Ota T., Suzuki Y.,  
RA Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,  
RA Isogai T., Sugano S.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AK027201; BAB15689.1; -  
DR HSSP; P15848; IFSU.  
SQ SEQUENCE 169 AA; 18905 MW; 4583C92FFFC3380 CRC64;

Query Match 11.5%; Score 63; DB 4; Length 169;  
Best Local Similarity 100.0%; Pred. No. 4.7e-57;  
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 488 STGKSVLNFNITADPYERVVLSNRPYGVKLLRLSQFNKTAVPVRYPPKDRSNPRLN 547  
|||||

Db 58 STGKSVLNFNITADPYERVVLSNRPYGVKLLRLSQFNKTAVPVRYPPKDRSNPRLN 117  
|||||

Qy 548 GGV 550

|||

Db 118 GGV 120

RESULT 2  
ID Q9NJU7 PRELIMINARY; PRT; 266 AA.  
AC Q9NJU7; 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE SULFATASE 2 (FRAGMENT).  
GN SULF2.  
OS Helix pomatia (Roman snail) (Edible snail).  
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Pulmonata; Stylommatophora;  
OC Helicaceae; Helicidae; Helix.  
OX NCBI\_TaxID=6536;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20233298; PubMed=10772344;  
RA Wittstock U., Fischer M., Svendsen I., Halkier B.A.;  
RT "Cloning and characterization of two cDNAs encoding sulfatases in the  
Roman snail, Helix pomatia.";  
RL IUBMB Life 49:71-76(2000).  
DR EMBL; AF109925; AAF30403.1; -.  
DR HSSP; P15848; IFSU.  
DR InterPro; IPR000917; Sulfatase.  
DR PROSITE; PS00523; SULFATASE\_1; 1.  
DR PROSITE; PS00149; SULFATASE\_2; 1.  
FT NON\_TER 266 266  
SQ SEQUENCE 266 AA; 30391 MW; 2F877BB1737A9C49 CRC64;  
  
Query Match 3.6%; Score 20; DB 5; Length 266;  
Best Local Similarity 100.0%; Pred. No. 4.6e-12;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 103 LDKLAAGVKLENYVOPIC 122  
Db 55 LDKLAAGVKLENYVOPIC 74  
|||||  
  
RESULT 3  
ID Q9NJU8 PRELIMINARY; PRT; 503 AA.  
AC Q9NJU8; 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE SULFATASE 1.  
GN SULF1.  
OS Helix pomatia (Roman snail) (Edible snail).  
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Stylommatophora;  
OC Helicaceae; Helicidae; Helix.  
OX NCBI\_TaxID=6536;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20233298; PubMed=10772344;  
RA Wittstock U., Fischer M., Svendsen I., Halkier B.A.;  
RT "Cloning and characterization of two cDNAs encoding sulfatases in the  
Roman snail, Helix pomatia.";  
RL IUBMB Life 49:71-76(2000).  
DR EMBL; AF109924; AAF30402.1; -.  
DR HSSP; P15848; IFSU.  
DR InterPro; IPR000917; Sulfatase.  
DR Pfam; PF00884; Sulfatase; 1.  
DR PROSITE; PS00149; SULFATASE\_2; 1.  
SQ SEQUENCE 503 AA; 55821 MW; AC45334BA694413D CRC64;  
  
Query Match 2.2%; Score 12; DB 5; Length 503;  
Best Local Similarity 100.0%; Pred. No. 0.0017;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 71 LENYVQPICTP 82  
  
RESULT 4  
ID Q9VQ95 PRELIMINARY; PRT; 217 AA.  
AC Q9VQ95; 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE CGI7006 PROTEIN.  
GN CGI7006.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=BERKELEY;  
MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
Ananidis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,  
Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
Bailey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
Dodonson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,  
Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
DR EMBL; AE003583; AAF51282.1; -.  
DR FlyBase; FBgn0031404; CGI7006.  
DR InterPro; IPR001604; Endonuclease.  
DR SMART; SM00477; NUC; 1.  
SQ SEQUENCE 217 AA; 24439 MW; ECABDBF82A5533C2 CRC64;

Query Match 1.5%; Score 8; DB 5; Length 217;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 504 ERVLSNR 511  
Db 35 ERVLSNR 42  
|||||

RESULT 5  
 Q9HYR2 PRELIMINARY; PRT; 330 AA.  
 AC Q9HYR2; 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE 3-OXOACYL-ACYL-CARRIER-PROTEIN.  
 GN FABH2 OR PA3333  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltzy L., Tolentino B., Westbrock-Wadman S., Yuan Y.,  
 RA Brody L.I., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 RT opportunistic pathogen.";  
 RL Nature 406:959-964(2000).  
 DR EMBL: AS004755; AAG06721.1; -.  
 DR HSSP: P24249; IEBL.  
 DR InterPro: IPR002106; AA\_trna\_ligase\_II.  
 DR PROSITE; PS00339; AA\_TRNA\_LIGASE\_II\_2; UNKNOWN\_1.  
 KW Complete proteome.  
 SQ SEQUENCE 330 AA; 34003 MW; 6F032604DF695F1E CRC64;

Query Match 1.5%; Score 8; DB 16; Length 330;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 EEGALLA 60  
 Db 171 EEGALLA 178  
 |||||

RESULT 6  
 Q92EX3 PRELIMINARY; PRT; 381 AA.  
 AC Q92EX3; 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE LIN0335 PROTEIN.  
 GN LIN0335.  
 OS Listeria innocua.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Listeria.  
 OX NCBI\_TaxID=1642;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CLIP 11262 / SEROVAR 6A;  
 RX PubMed=11679669;  
 RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,  
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,  
 RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,  
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,  
 RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,  
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,  
 RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkut G.,  
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,  
 RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,  
 RA Remmel B., Rose M., Schlueder T., Simoes N., Tierrez A.,  
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;

RT "Comparative genomics of Listeria species.";  
 RL Science 294:849-852(2001).  
 DR EMBL: AL596164; CAC95568.1; -.  
 DR ListList; LIN00335; -.  
 KW Complete proteome.  
 SQ SEQUENCE 381 AA; 43625 MW; D74B4FEE1C1A720C CRC64;

Query Match 1.5%; Score 8; DB 16; Length 381;  
 Best Local Similarity 100.0%; Pred. No. 20;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 315 GFYNNII 322  
 Db 181 GFYNNII 188  
 |||||

RESULT 7  
 Q18924 PRELIMINARY; PRT; 452 AA.  
 AC Q18924; 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE HYPOTHETICAL 51.7 KDA PROTEIN.  
 GN D1014.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA du Z., Leimbac D.;  
 RT "The sequence of C. elegans cosmid D1014.";  
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Waterston R.;

Query Match 1.5%; Score 8; DB 5; Length 452;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 MVGKWHLG 180  
 Db 134 MVGKWHLG 141  
 |||||

RESULT 8  
 Q9CKEO PRELIMINARY; PRT; 453 AA.  
 ID Q9CKEO  
 AC Q9CKEO;

```

DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN PM1682.
GN PM1682.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70."
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AE006204; AK03766.1; -.
DR InterPro; IPR000566; Lipocin_cyFABP.
DR Pfam; PF00884; Sulfatase; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 453 AA; 51449 MW; 83B91532C6252B0E CRC64;

Query Match 1.5%; Score 8; DB 16; Length 453;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 100 TPTLDKLA 107
Db 29 TPTLDKLA 36
|||||||

RESULT 9
Q9V6F2 PRELIMINARY; PRT; 542 AA.
AC Q9V6F2;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE C86646 PROTEIN.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaiswal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

```

```

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleeb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003821; AAF58475.1; -.
DR HSSP; PI5848; IFSU.
DR FlyBase; FBgn0033763; CG8646.
DR InterPro; IPR000917; Sulfatase.
DR Pfam; PF00884; Sulfatase; 1.
DR PROSITE; PS00523; SULFATASE_1; 1.
DR PROSITE; PS00149; SULFATASE_2; 1.
SQ SEQUENCE 542 AA; 60278 MW; EACD9C14DD2B3204 CRC64;

Query Match 1.5%; Score 8; DB 5; Length 542;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 120 PICTPSRS 127
Db 71 PICTPSRS 78
|||||||

RESULT 10
Q25384 PRELIMINARY; PRT; 551 AA.
ID Q25384;
AC Q25384;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE ARYL SULFATASE.
GN ARS.
OS Hemichordatus pulcherrimus (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
OC Hemichordatus.
OX NCBI_TaxID=7650;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPERM;
RA Yamada K.;
RL Submitted (SEP-1989) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE OF 1-105 FROM N.A.
RC TISSUE=SPERM;
RX MEDLINE=900921130; PubMed=2598936;
RA Yamada K., Akasaka K., Shimada H.;
RT "Structure of sea-urchin arylsulfatase gene."
RL Eur. J. Biochem. 186:405-410(1989).
[3]
RP SEQUENCE FROM N.A.
RC TISSUE=SPERM;
RA Shimada H.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE FROM N.A.
RC TISSUE=SPERM;
RA Akasaka K., Sakamoto N., Yamamoto T., Morokuma J., Fujikawa N.,
RA Takata K., Eguchi S., Shimada H.;
RT "Corrected structure of the 5' flanking region of arylsulfatase gene of

```



RT the sea urchin, Hemicentrotus pulcherrimus.\*;  
 RL Dev. Growth Differ. 36:633-636(1994).  
 DR EMBL; X16679; CAA34667.1; .  
 DR EMBL; X16680; CAA34667.1; JOINED.  
 DR EMBL; X16681; CAA34667.1; JOINED.  
 DR EMBL; X16682; CAA34667.1; JOINED.  
 DR EMBL; X16683; CAA34667.1; JOINED.  
 DR EMBL; X16684; CAA34667.1; JOINED.  
 DR HSSP; P15289; IAUk.  
 DR InterPro; IPR000917; Sulfatase.  
 DR Pfam; PF00884; Sulfatase; 1.  
 DR PROSITE; PS00523; SULFATASE\_1; 1.  
 DR PROSITE; PS00149; SULFATASE\_2; 1.  
 SQ SEQUENCE 551 AA; 60943 MW; 544315A4880E12C1 CRC64;

Query Match 1.5%; Score 8; DB 5; Length 551;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 MVGKWHLG 180

Db 153 MVGKWHLG 160

RESULT 11

O16138 PRELIMINARY; PRT; 559 AA.  
 AC O16138;

DT 01-JAN-1998 (TrEMBLrel. 05, Created)

DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE ARYL SULFATASE.

GN ARS.

OS Helicoidaris erythrogramma (Sea urchin).

OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;

OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Echinometridae;

OC Helicoidaris.

OX NCBI\_TaxID=7634;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98298333; PubMed=9634485;

RA Haag E.S., Raff R.A.;

RT "Isolation and characterization of three mRNAs enriched in embryos of

the direct-developing sea urchin *Helicoidaris erythrogramma*: evolution

of larval ectoderm.";

RL Dev. Genes Evol. 208:188-204(1998).

RL EMBL; AF013158; AAC27821.1; .

DR HSSP; P15289; IAUk.

DR InterPro; IPR000917; Sulfatase.

DR Pfam; PF00884; Sulfatase; 1.

DR PROSITE; PS00523; SULFATASE\_1; 1.

DR PROSITE; PS00149; SULFATASE\_2; 1.

SQ SEQUENCE 559 AA; 61915 MW; 13DE7DCEBD05752 CRC64;

Query Match 1.5%; Score 8; DB 5; Length 559;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 MVGKWHLG 180

Db 159 MVGKWHLG 166

RESULT 12

Q9VVM4

ID Q9VVM4 PRELIMINARY; PRT; 579 AA.

AC Q9VVM4;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE CG7402 PROTEIN.

GN CG7402.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt J.G., Nelson C.R., Miklos D.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M., Lai Z.,  
 RA Palazzolo M., Palatnik G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Fang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RL EMBL; AE003522; AAF49287.1; .  
 DR HSSP; P15848; 1FSU.  
 DR FlyBase; FBgn0036768; CG7402.  
 DR InterPro; IPR000917; Sulfatase.  
 DR Pfam; PF00884; Sulfatase; 1.  
 DR PROSITE; PS00149; SULFATASE\_2; 1.  
 SQ SEQUENCE 579 AA; 65053 MW; EBF2F4F13D759298 CRC64;

Query Match 1.5%; Score 8; DB 5; Length 579;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 VGRKHLGF 181

Db 127 VGRKHLGF 134

RESULT 13

Q93D92

ID Q93D92 PRELIMINARY; PRT; 898 AA.

AC Q93D92;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

```
DE PSAB.
GN PSAB.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=L711;
RA Tao L., Tanzer J.M.;
RT "Novel sucrose-dependent adhesion cofactors (sdc) in Streptococcus
RT mutans.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF397166; AAL04089.1; -
SQ SEQUENCE 898 AA; 101903 MW; A2D9CA6GF9357AF11 CRC64;

Query Match 1.5%; Score 8; DB 2; Length 898;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 195 TFFGSLLC 202
Db 440 TFFGSLLC 447
|||||||

RESULT 14
Q9VVK9 PRELIMINARY; PRT; 996 AA.
AC Q9VVK9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CG5584 PROTEIN.
GN CG5584.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
```

```
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003522; AAP49302.1; -
DR HSP; P15848; IfSU.
DR FlyBase; FBgn0036755; CG5584.
DR InterPro; IPR000917; Sulfatase.
DR Pfam; PF00884; Sulfatase; 1.
DR PROSITE; PS00149; SULFATASE-2; 1.
SQ SEQUENCE 996 AA; 113514 MW; C92AA378D082FB3 CRC64;

Query Match 1.5%; Score 8; DB 5; Length 996;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 174 VGKWLGF 181
Db 126 VGKWLGF 133
|||||||

RESULT 15
Q9AXB1 PRELIMINARY; PRT; 1002 AA.
ID Q9AXB1;
AC Q9AXB1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE P0501G01.6 PROTEIN.
GN P0501G01.6.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0501G01.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP002819; BAB21077.1; -
DR InterPro; IPR001871; bZIP.
DR InterPro; IPR001296; Glycos.transf_1.
DR Pfam; PF00534; Glycos.transf_1; 1.
DR PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
SQ SEQUENCE 1002 AA; 113136 MW; F459312DD89CC5EF CRC64;

Query Match 1.5%; Score 8; DB 10; Length 1002;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 GHPPPPSP 15
Db 24 GHPPPPSP 31
|||||||
```

Search completed: July 23, 2002, 19:09:59  
Job time: 194 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 19:02:54 ; Search time 38.31 Seconds  
(without alignments)  
1594.638 Million cell updates/sec

Title: US-09-495-823-7  
Perfect score: 550  
Sequence: 1 MAPRCAGHPPPSPQACVC.....VPRYPKDRSPNRLNGGV 550

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 0

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A\_Geneseq\_032802.\*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	385	70.0	599	22 AAB85483	Human 23553 sulfat
2	289	52.5	289	22 AAU14421	Human novel protei
3	196	35.6	22	AAU14185	Human novel protei
4	181	32.9	515	20 AAY41701	Human PRO708 prote
5	181	32.9	515	21 AAB44257	Human PRO708 (UNQ3
6	181	32.9	515	22 AAU29061	Human PRO polypept
7	179	32.5	187	22 AAM25415	Human protein sequ
8	109	19.8	128	22 AAU23335	Novel human enzyme
9	19	3.5	569	22 AAB85482	Human 25278 sulfat
10	10	1.8	533	22 AAB51184	Human sulfatase pr
11	8	1.5	217	22 ABB66123	Drosophila melanog

12	8	1.5	330	22 AAU36321	Pseudomonas aerugi
13	8	1.5	542	22 ABB63962	Drosophila melanog
14	8	1.5	551	22 AAB51186	Sea urchin sulfata
15	8	1.5	579	22 ABB63246	Drosophila melanog
16	8	1.5	996	22 ABB61654	Drosophila melanog
17	8	1.5	1420	18 AAW13826	Yeast transcriptio
18	7	1.3	15	22 AAB86555	Human cytomagalovi
19	7	1.3	15	22 AAB86556	Human cytomagalovi
20	7	1.3	15	22 AAB86557	Human cytomagalovi
21	7	1.3	16	22 AAB86557	Human cytomagalovi
22	7	1.3	16	22 AAB86557	HIV-1 Tat peptide
23	7	1.3	19	21 AAB10389	HPV E6 protein inh
24	7	1.3	28	14 AAR39867	E2 peptide E2-15,
25	7	1.3	34	22 AAO12394	Human polypeptide
26	7	1.3	45	22 AAM23940	Human EST encoded
27	7	1.3	48	22 ABB70089	Drosophila melanog
28	7	1.3	51	22 AAM83802	Human immune/haema
29	7	1.3	55	22 AAU22229	Human cardiovascular
30	7	1.3	62	22 ABB27456	Human peptide #107
31	7	1.3	62	22 ABB32605	Peptide #111 encod
32	7	1.3	62	22 ABB18101	Protein #100 encod
33	7	1.3	62	22 AAM53434	Human brain expres
34	7	1.3	62	22 AAM65812	Human bone marrow
35	7	1.3	62	22 AAM13674	Peptide #108 encod
36	7	1.3	62	22 AAM26073	Peptide #110 encod
37	7	1.3	62	22 AAM01424	Peptide #106 encod
38	7	1.3	75	22 AAU29568	Novel human secret
39	7	1.3	79	22 AAG01102	Novel human diagno
40	7	1.3	85	22 AAU29804	Novel human secret
41	7	1.3	86	22 AAB46684	HIV-1 Tat peptide
42	7	1.3	105	22 AAO06154	Human polypeptide
43	7	1.3	110	22 AAU31767	Novel human secret
44	7	1.3	112	22 AAG06329	Novel human diagno
45	7	1.3	113	22 AAG01780	Human polypeptide
			115	21 AAG57479	Arabidopsis thalia

ALIGNMENTS

RESULT 1	
AA885483	
ID	AA885483 standard; Protein; 599 AA.
AC	AA885483;
XX	25-SEP-2001 (first entry)
DT	Human 23553 sulfatase polypeptide.
XX	Sulfatase; 23553 sulfatase; nootropic; neuroprotective; antibacterial;
KW	antiinflammatory; vasotropic; antitumor; gene therapy; human.
XX	Homo sapiens.
OS	WO200155411-A2.
PN	02-AUG-2001.
XX	31-JAN-2001; 2001WO-US03266.
PF	31-JAN-2000; 2000US-0495823.
XX	(MILL-) MILLENNIUM PHARM INC.
XX	Glucksmann MA, Williamson M, Rudolph-Owen LA, Tsai F;
PI	WPI; 2001-476214/51.
XX	N-PSDB; AAH46863, AAH46867.
DR	Novel human sulfatase polypeptides useful for treating and diagnosing
XX	sulfatase-related disorders such as cerebrovascular diseases, acute
PT	meningitis, multiple sclerosis, degenerative diseases and tumor -
XX	

PS Claim 9; Fig 15; 180pp; English.

XX The invention provides 22438, 23553, 25278 or 26212 human sulfatase

CC polypeptides and polynucleotides. The sulfatase genes and polypeptides

CC are useful for treating disorders involving the brain such as

CC cerebrovascular diseases, infections such as acute meningitis,

CC demyelinating diseases including multiple sclerosis, degenerative

CC diseases affecting the cerebral cortex including Alzheimer's disease

CC and Pick disease, spinocerebellar degenerations including spinocerebellar

CC ataxias including Friedreich ataxia, and ataxia telangiectasia,

CC degenerative diseases affecting motor neurons including amyotrophic

CC lateral sclerosis, inborn errors of metabolism such as leukodystrophies,

CC toxic and acquired metabolic diseases, including vitamin deficiencies,

CC and neurocutaneous syndromes (phakamatoses) including neurofibromatosis.

CC The present sequence represents a human 23553 sulfatase polypeptide.

XX SQ Sequence 599 AA;

Query Match 70.0%; Score 385; DB 22; Length 599;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 485; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MAPGCGAGHPPPSPQACVCPGKMLAGFWILCLLTGYLSWGQALEEEEGALLA 60

DB 1 maprgcaghhppspgacvcpqkmlangalagfwilclltgyylswgaleeeegalla 60

QY 61 QAGCKLEPSTTSQPHLIILADDDQFROVGHGSEIKTPTLDKLAAGVKLENYVQP 120

DB 61 qagcklepsttsqphliiladdqgfrdvghgseiktptldklaegvklennyvqp 120

QY 121 ICTPSRQFITGKYQIHTGLQHSIIRTPQNCPLDNLATPQKLKEGYSTHMGKWHLG 180

DB 121 ictpsrqfitgkyqihgtlqhsirtpqncpldnlattpqlkkgysthmvgkwhlg 180

QY 181 FYRKECMPTRRGFTFGSLGSDGYTHYKCDSPGCMGYDLYENDNAWDYDNGIYSTQ 240

DB 181 fyrkecmptrrgftfgslgsdgythykcdspgcmgydlyendnaawdydngiystq 240

QY 241 MYTORVQOILASHNPTKPIFYLTAYQAVHSPLOAPGRYFEHYRSIINRRYAAMLSCL 300

DB 241 mytrvqqlashnptkpfilytayqavhsploapgrfyfhyrsilnrrryaamlsc 300

QY 301 DEAINNVTALKTYGFYNNSSIIYSSDNGGQPTAGGSNWPLRSGKTYWEGGIRAVGFVH 360

DB 301 deainnvtalktygfynssiiyssdnggqptaggsnwplrgsktyweggiravgf 360

QY 361 SPLKNGKTVCKEPVHTIDWYPTLISLAEGQIDEDIQDGYDIWETISEGLRSPRVILH 420

DB 361 spllknkgtvckelvhitdwypthlislaegqidediqldgydiwetiseglrsprv 420

QY 421 NIDPIYTKAKNGSWAAGYGIWNTAIQSAIRVQHWKLLTGNPGYSDWVPQPSFSLGNRW 480

DB 421 nidpiytkakngswaagygiwntaisairvqhkwlltgnpgysdwvppqpsfslgn 480

QY 481 HNERIT 486

DB 481 hnerit 486

RESULT 2

ID AAU14421

XX AAU14421 standard; Protein; 289 AA.

AC AAU14421;

XX 24-OCT-2001 (first entry)

DE Human novel protein #292.

KW Human; novel protein; Antianaemic; osteopathic; antiinflammatory;

KW immunomodulatory; cytostatic; neuroprotective; vulnerable; nootropic;

KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;

KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;

KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;

KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;

XX tissue regeneration; immune disorder.

OS Homo sapiens.

PN WO200155437-A2.

XX 02-AUG-2001.

XX 25-JAN-2001; 2001WO-US02623.

XX 25-JAN-2000; 2000US-0491404.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

XX WPI: 2001-451939/48.

DR N-PSDB; AAS22726.

XX Isolated polypeptides useful for treating anti-inflammatory diseases,

PT nervous system disorders, and for regenerating bone and cartilage -

XX Example 4; Page 809-810; 894pp; English.

XX The invention relates to polynucleotides encoding novel human

CC proteins or their active domains. The polypeptides, polynucleotides and

CC antibodies raised against the polypeptides are used in a method of

CC treatment of a mammal and prevention of disorders caused by the aberrant

CC protein expression or activity. The polypeptides can be used as

CC molecular weight markers, food supplements, and in antibody production.

CC The polypeptides are used to identify compounds which bind to the

CC polypeptides. Polynucleotides of the invention are used as probes and

CC primers, for sequencing, for chromosome or gene mapping, in the

CC production of recombinant proteins, and in generating anti-sense DNA or

CC RNA and in gene therapy. Polypeptides of the invention can be used to

CC target drugs to a tumour, in assays to determine biological activity, to

CC raise antibodies/ elicit an immune response, to determine quantitative

CC protein levels, as tissue markers, and to isolate receptors or ligands.

CC Polypeptides of the invention may also be useful in treating platelet

CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,

CC ligament and/or nerve tissue, wound healing, treating burns, promoting

CC the proliferation, differentiation and survival of stem cells, as a

CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,

CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral

CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or

CC fungal infection or from autoimmunity, cancer, allergy, asthma,

CC graft-versus-host disease, eczema, haemophilia, thrombosis,

CC anti-inflammatory diseases, nervous system disorders, and infection.

XX CC The present sequence represents a protein of the invention.

XX SQ Sequence 289 AA;

Query Match 52.5%; Score 289; DB 22; Length 289;

Best Local Similarity 100.0%; Pred. No. 2.4e-285;

Matches 289; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 MLAMGALAGFWILCLLTGYLSWGQALEEEEGALLAQAGEKLEPSTTSQPHLIIFILA 83

DB 1 mlamgalagfwilclltgyylswgaleeeegallaqageklepsttsqphliffila 60

QY 84 DDQCFRDVGHGSEIKTPTLDKLAAGVKLENYVQPICTPSRQFITGKYQIHTGLQHS 143

DB 61 ddqgfrdvghgseiktptldklaegvklennyvqpictpsrqfitgkyqihgtlqhs 120

QY 144 IIRPTQNCPLDNLATPQKLKEGYSTHMGKWHLGFRYRKECMPTRRGFTFGSLGSG 203

DB 121 iirptqncpldnlattpqlkkgysthmvgkwhlgfyrkecmptrrgftfgslgsg 180

QY 204 GDYTYHYKCDSPGCMGYDLYENDNAWDYDNGIYSTQMYTORVQOILASHNPTKPIFYI 263

Db 181 gdytyhkydsgpncgydlyendaawdydngiydstqmytqrvcqilashnptkpfilyi 240  
Qy 264 AYQAVHSPLOAPGRYFHEHRSIIINRRRYAAMLSCIDEAINNVTLALK 312  
Db 241 ayqavhsplqapgrfyfhehrysiinihrrryaamiscideainnvtlalk 289  
RESULT 3  
AAU14185  
ID AAU14185 standard; Protein; 366 AA.  
XX  
AC AAU14185;  
XX  
DT 24-OCT-2001 (first entry)  
XX  
DE Human novel protein #56.  
XX  
KW Human; novel protein; Antianaemic; osteopathic; antinflammatory;  
KW immunomodulatory; cytostatic; neuroprotective; vulnery; nootropic;  
KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;  
KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;  
KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;  
KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;  
KW tissue regeneration; immune disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200155437-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 25-JAN-2001; 2001WO-US02623.  
XX  
PR 25-JAN-2000; 2000US-0491404.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT;  
XX  
DR WPI: 2001-451939/48.  
DR N-PSDB: AAS22490.  
XX  
Isolated polypeptides useful for treating anti-inflammatory diseases,  
nervous system disorders, and for regenerating bone and cartilage -  
XX  
PS Example 4; Page 557-558; 894pp; English.  
XX  
CC The invention relates to polynucleotides encoding novel human  
CC proteins or their active domains. The polypeptides, polynucleotides and  
CC antibodies raised against the polypeptides are used in a method of  
CC treatment of a mammal and prevention of disorders caused by the aberrant  
CC protein expression or activity. The polypeptides can be used as  
CC molecular weight markers, food supplements, and in antibody production.  
CC The polypeptides are used to identify compounds which bind to the  
CC polypeptides. Polynucleotides of the invention are used as probes and  
CC primers, for sequencing, for chromosome or gene mapping, in the  
CC production of recombinant proteins, and in generating anti-sense DNA or  
CC RNA and in gene therapy. Polypeptides of the invention can be used to  
CC target drugs to a tumour, in assays to determine biological activity, to  
CC raise antibodies/elicit an immune response, to determine quantitative  
CC protein levels, as tissue markers, and to isolate receptors or ligands.  
CC Polypeptides of the invention may also be useful in treating platelet  
CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,  
CC ligament and/or nerve tissue, wound healing, treating burns, promoting  
CC the proliferation, differentiation and survival of stem cells, as a  
CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,  
CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral  
CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or  
CC fungal infection or from autoimmunity, cancer, allergy, asthma;  
CC graft-versus-host disease, eczema, haemophilia, thrombosis,  
CC anti-inflammatory diseases, nervous system disorders, and infection.  
CC The present sequence represents a protein of the invention.

XX  
SQ Sequence 366 AA;  
Query Match 35.6%; Score 196; DB 22; Length 366;  
Best Local Similarity 100.0%; Pred. No. 1.3e-190;  
Matches 196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 16 QACVCPGKMLAMGALAGFWILCLLTGYLSWGQALEEEBEGALLAQAGEKLEPSTTSQ 75  
Db 162 qacvcpgkmlamgalagfwilclltygylswgqaleeeegallaqageklepsttsq 221  
Qy 76 PHLIFILADQGFDRVGVYHGSEIKTPTLDKLAABGVKLENYVQPICTPSRSQFITGKYQ 135  
Db 222 phliladddgfrdvgyhgseiktptldklaaegvklennyvqpictpsrsqfitgkyq 281  
Qy 136 IHTGLQHSIIIRPTQPNCLPLDNATLPQKLKEVGYSTHVMGKWHLGFYRKECMPTRRGDT 195  
Db 282 ihtglqhsilrptqpnclpldnatlpqklkevgysthvmgkwhlgfyркеcmptrrgdt 341  
Qy 196 PFGSLGSGDYTHYK 211  
Db 342 ffgslgsgdythyk 357  
RESULT 4  
AA41701  
ID AAY41701 standard; Protein; 515 AA.  
XX  
AC AAY41701;  
XX  
DT 07-DEC-1999 (first entry)  
XX  
DE Human PRO708 protein sequence.  
XX  
KW Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;  
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;  
KW secreted protein; transmembrane protein.  
XX  
OS Homo sapiens.  
XX  
PN WO9946281-A2.  
XX  
PD 16-SEP-1999.  
XX  
PF 08-MAR-1999; 99WO-US05028.  
XX  
PR 10-MAR-1998; 98US-0077450.  
PR 11-MAR-1998; 98US-0077632.  
PR 11-MAR-1998; 98US-0077641.  
PR 11-MAR-1998; 98US-0077649.  
PR 12-MAR-1998; 98US-0077791.  
PR 13-MAR-1998; 98US-0078004.  
PR 17-MAR-1998; 98US-0040220.  
PR 20-MAR-1998; 98US-0078886.  
PR 20-MAR-1998; 98US-0078910.  
PR 20-MAR-1998; 98US-0078936.  
PR 20-MAR-1998; 98US-0078939.  
PR 25-MAR-1998; 98US-0079294.  
PR 26-MAR-1998; 98US-0079656.  
PR 27-MAR-1998; 98US-0079663.  
PR 27-MAR-1998; 98US-0079664.  
PR 27-MAR-1998; 98US-0079689.  
PR 27-MAR-1998; 98US-0079728.  
PR 27-MAR-1998; 98US-0079786.  
PR 30-MAR-1998; 98US-0079920.  
PR 30-MAR-1998; 98US-0079923.  
PR 31-MAR-1998; 98US-0080105.  
PR 31-MAR-1998; 98US-0080107.  
PR 31-MAR-1998; 98US-0080165.  
PR 31-MAR-1998; 98US-0080194.  
PR 01-APR-1998; 98US-0080327.  
PR 01-APR-1998; 98US-0080328.







CC the proliferation or differentiation of chondrocyte cells. The PRO  
CC proteins can be used to determine the presence of tumours and also  
CC susceptibility to tumour development, particularly adrenal, lung, colon,  
CC breast, prostate, rectal, cervical, or liver tumours, in mammalian  
CC subjects. The oligonucleotide probes specific for the PRO nucleic acids  
CC can be used for genetic analysis of individuals with genetic disorders.  
XX  
SQ Sequence 515 AA;  
  
Query Match 32.9%; Score 181; DB 22; Length 515;  
Best Local Similarity 100.0%; Pred. NO. 3.4e-175;  
Matches 181; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 MAPRCAGHPPPPSPQACVCPGKMLAMGALAGFWILCLLTGYLSWGOALREEGALLA 60  
Db 1 mapr9caghp9ppspqacvcp9gkmlangalagfwilclltygylswgqaleeeegalla 60  
|||||  
Qy 61 QAGEKLEPSTTSQPHLFIADDDQGRDVGYPHGSEIKTPTLDKLAEGVKLENYVQP 120  
Db 61 qageklepsttsqphlfiiladdqgrdvgvghgseikptldklaaegvklennyvqp 120  
|||||  
Qy 121 ICTPSRSQFITGKQIHTGLQHSIIRTPQNCPLDNPATLPQKLEKGVSTHYMKVGHG 180  
Db 121 ictpsrsqfitgkqihgtglqhsirtpqncpldnpatlpqklevgsthykmvghg 180  
|||||  
Qy 181 F 181  
Db 181 f 181  
  
RESULT 7  
AAU25415  
ID AAU25415 standard; Protein; 187 AA.  
XX  
AC AAU25415;  
XX  
DT 16-OCT-2001 (first entry)  
XX  
DE Human protein sequence SEQ ID NO:930.  
XX  
KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
KW anti-inflammatory; antirheumatic; antiarthritic; immunosuppressive;  
KW antibacterial; endocrine; cardiac; central nervous system; virucide;  
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
KW antiagregant; haemostatic; vulnerary; antiulcer; osteopathic; eczema;  
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;  
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
KW neurological disorder.  
XX  
OS Homo sapiens.  
XX  
XX WO200153455-A2.  
XX  
XX 26-JUL-2001.  
XX  
XX 22-DEC-2000; 2000WO-US35017.  
XX  
XX 23-DEC-1999; 99US-0471275.  
XX 21-JAN-2000; 2000US-0488725.  
XX 25-APR-2000; 2000US-0552317.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT;  
XX

DR WPI; 2001-457603/49.  
DR N-PSDB; AAH99356.  
XX  
PT Isolated human polynucleotides encoding polypeptides. useful for the  
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -  
XX  
PS Claim 20; Page 200; 1217pp; English.  
XX  
CC AAH99166 to AAH99904 encode the human proteins given in AAU25225 to  
CC AAU25963. The proteins can have activities based on the tissues and  
CC cells they are expressed in, such as: antiinflammatory; antirheumatic;  
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;  
CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;  
CC cardiovascular; antianaemic; antiagregant; haemostatic; vulnerary;  
CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;  
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;  
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
CC encoding them can be used in gene therapy, antisense therapy and vaccine  
CC production. The proteins and polynucleotides are useful for screening for  
CC agonists or antagonists of a protein and for the treatment and diagnosis  
CC of disorders associated with the activity of a protein e.g. inflammation,  
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,  
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,  
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,  
CC osteoporosis, severe combined immunodeficiency, eczema, allergic  
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and  
CC neurological disorders.  
XX  
SQ Sequence 187 AA;  
  
Query Match 32.5%; Score 179; DB 22; Length 187;  
Best Local Similarity 100.0%; Pred. NO. 1.5e-173;  
Matches 179; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 134 YQIHTGLQHSIIRTPQNCPLDNPATLPQKLEKGVSTHYMKVGHGKWHLGFKYRKECMPTRRGF 193  
Db 9 yqihgtglqhsirtpqncpldnpatlpqklevgsthykmvghlgfyrkecmptrrgf 68  
|||||  
Qy 194 DTFEGSLGSGDYTHYKCDSPGCMGYDLYENDNAWDYDNGIYSTQMYTORVQOILASH 253  
Db 69 dtfgslgsgdytyhykcdspgcmgydlyendnaawdydnglystqmytrvqqilash 128  
|||||  
Qy 254 NPTKPIFLYIAYQAVHSPLOAPGVFEHYRSIINIRRYAAMLSCLDENNVTLAK 312  
Db 129 nptkpiflyiayqavhsploapgvfeyhrysiinirryaamlscldeainnvtlak 187  
|||||  
  
RESULT 8  
AAU23335  
ID AAU23335 standard; Protein; 128 AA.  
XX  
AC AAU23335;  
XX  
DT 17-DEC-2001 (first entry)  
XX  
DE Novel human enzyme polypeptide #421.  
XX  
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;  
KW ligase; hyperproliferative disorder; immunodeficiency disorder;  
KW autoimmune disorder; neurological disorder; metabolic disorder;  
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;  
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;  
KW nephrotropic; anticoagulant.  
XX  
OS Homo sapiens.  
XX  
XX WO200155301-A2.  
XX  
XX 02-AUG-2001.  
XX

PF 17-JAN-2001; 2001WO-US01239.  
XX  
31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0232403.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0234998.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246533.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX Rosen CA, Barash SC, Ruben SM;  
PI  
XX WPI; 2001-465566/50.  
XX N-PSDB; AAS41205.  
XX

PT Novel polypeptides and polynucleotides useful for diagnosing,  
PT preventing, treating neural, immune system, muscular, reproductive,  
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous  
PT diseases  
XX  
PS Claim 11; SEQ ID No 1331; 1180pp; English.  
XX  
CC The present invention relates to the isolation of novel human enzyme  
CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences  
CC encoding them. The enzyme polypeptides of the invention may comprise the  
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,  
CC isomerases or ligases. The sequences of the invention are useful in the  
CC diagnosis, treatment, prevention and/or prognosis of a wide range of  
CC disorders including hyperproliferative disorders (e.g. cancer),  
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders  
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),  
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders  
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),  
CC blood-related disorders (e.g. haemophilia), reproductive disorders  
CC (e.g. infertility) and infectious disorders (e.g. Influenza). The  
CC polynucleotides of the invention can also be used in gene therapy.  
CC AAU22915-AAU23814 represent the novel human enzyme polypeptides of the  
CC invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 128 AA;

Query Match 19.8%; Score 109; DB 22; Length 128;  
Best Local Similarity 100.0%; Pred. No. 1.8e-102;  
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 90 DVGTHGSEIKPTLDKLAAGVCKLENYVQIPCTPSPRSQFTGKYQHTGQLQHSIIIRPTQ 149  
Db 6 dvgyhgseiktptldklaaegvcklenyvvqipctpsrsqftgkyqhtgqlqhsilrptq 65  
QY 150 PNCPLDNPATLPQKLKEVGSTHMGVKGWHLGFRKCEMPTRRGDTFFG 198  
Db 66 pncpldnpatlpqklkevgsthmvgkwhlgfrkcekmptrrgdtffg 114

RESULT 9  
AAB85482  
ID AAB85482 standard; Protein; 569 AA.  
XX  
AC AAB85482;  
XX  
DT 25-SEP-2001 (first entry)  
XX  
DE Human 25278 sulfatase polypeptide.  
XX  
KW Sulfatase; 25278 sulfatase; nootropic; neuroprotective; antibacterial;  
KW antiinflammatory; vasotropic; antitumor; gene therapy; human.  
XX  
OS Homo sapiens.  
XX  
PN W0200155411-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 31-JAN-2001; 2001WO-US03266.  
XX  
PR 31-JAN-2000; 2000US-0495823.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
XX  
PI Glucksmann MA, Williamson M, Rudolph-Owen LA, Tsai F;  
XX WPI; 2001-476214/51.  
DR N-PSDB; AAH46862, AAH46866.  
XX

PT Novel human sulfatase polypeptides useful for treating and diagnosing  
PT sulfatase-related disorders such as cerebrovascular diseases, acute  
PT meningitis, multiple sclerosis, degenerative diseases and tumor  
XX  
PS Claim 9; Fig 10; 180pp; English.  
XX  
CC The invention provides 22438, 23553, 25278 or 26212 human sulfatase  
CC polypeptides and polynucleotides. The sulfatase genes and polypeptides  
CC are useful for treating disorders involving the brain such as  
CC cerebrovascular diseases, infections such as acute meningitis,  
CC demyelinating diseases including multiple sclerosis, degenerative  
CC diseases affecting the cerebral cortex including Alzheimer's disease  
CC and Pick disease, spinocerebellar degenerations including spinocerebellar  
CC ataxias including Friedreich ataxia, and ataxia telangiectasia,  
CC degenerative diseases affecting motor neurons including amyotrophic  
CC lateral sclerosis, inborn errors of metabolism such as leukodystrophies,  
CC toxic and acquired metabolic diseases, including vitamin deficiencies,  
CC and neurocutaneous syndromes (phakomatosis) including neurofibromatosis.  
CC The present sequence represents a human 25278 sulfatase polypeptide.  
XX  
SQ Sequence 569 AA;

Query Match 3.5%; Score 19; DB 22; Length 569;  
Best Local Similarity 100.0%; Pred. No. 2.7e-10;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 GYSTHMVGKWHLGFRKCE 186  
Db 139 gysthmvgkwhlgfrkcec 157

RESULT 10  
AAB51184  
ID AAB51184 standard; Protein; 533 AA.  
XX  
AC AAB51184;  
XX  
DT 21-MAR-2001 (first entry)  
XX  
DE Human sulfatase protein B SEQ ID NO:13.  
XX  
KW Human; iduronate 2-sulfatase; IDS; glycosylated; gene therapy;  
KW mucopolysaccharidosis inhibitor; IDS deficiency disorder;  
KW Hunter syndrome; mucopolysaccharidosis type II.  
XX  
OS Homo sapiens.  
XX  
PN US6153188-A.  
XX  
PD 28-NOV-2000.  
XX  
PF 12-FEB-1999; 99US-0249003.  
XX  
PR 17-DEC-1992; 92US-0991973.  
PR 28-NOV-1994; 94US-0345212.  
PR 12-NOV-1991; 91US-0790362.  
XX  
PA (WOME-) WOMEN'S & CHILDREN'S HOSPITAL.  
XX  
PI Wilson PJ, Anson DS, Occhiodoro T, Bielicki J, Clements PR;  
PI Hopwood JJ, Morris CP;  
XX  
DR WPI; 2001-060076/07.  
XX  
PT New highly glycosylated recombinant human iduronate 2-sulfatase (IDS)  
PT useful for diagnosing or treating subjects suspected of having or  
PT suffering from IDS deficiency disorders, e.g. Hunter syndrome  
PT (mucopolysaccharidosis-II) -  
XX  
PS Example 1; Fig 3; 53pp; English.  
XX  
CC The present invention describes a recombinant human iduronate 2-sulfatase

CC (IDS). The recombinant IDS is more highly glycosylated than the naturally  
 CC occurring enzyme isolated from human tissue. The recombinant human IDS  
 CC can be produced in Chinese Hamster Ovary (CHO) cells or in a human cell.  
 CC The recombinant IDS comprises a fusion protein. It is a  
 CC mucopolysaccharidosis inhibitor and can be used in gene therapy.  
 CC The recombinant IDS is useful in treating and diagnosing subjects  
 CC suffering from or suspected of having IDS deficiency disorders, e.g.  
 CC Hunter syndrome (mucopolysaccharidosis type II). The present sequence  
 CC represents a human sulfatase which is used in comparison with the  
 CC IDS sequence in an example from the present invention.

XX SO Sequence 533 AA;

Query Match 1.8%; Score 10; DB 22; Length 533;  
 Best Local Similarity 100.0%; Pred. No. 0.37; Length 533;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 171 THMVGKWHLG 180  
 |||||  
 Db 140 thmvgkwhlg 149

#### RESULT 11

ABB66123  
 ID ABB66123 standard; Protein; 217 AA.

AC ABB66123;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 25161.

KW Drosophila; developmental biology; cell signalling; insecticide;  
 pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

PI WPI; 2001-656860/75.

DR N-PSDB; ABL10226.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -

PS Disclosure; SEQ ID NO 25161; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SO Sequence 217 AA;

Query Match 1.5%; Score 8; DB 22; Length 217;  
 Best Local Similarity 100.0%; Pred. No. 18;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 504 ERVDSNR 511  
 |||||  
 Db 35 ervdlsnr 42

#### RESULT 12

AAU36321  
 ID AAU36321 standard; Protein; 330 AA.

XX AC AAU36321;

XX 14-FEB-2002 (first entry)

XX Pseudomonas aeruginosa cellular proliferation protein #311.

KW Antisense; prokaryotic cellular proliferation protein;  
 KW antibiotic; antibacterial; drug design.

XX OS Pseudomonas aeruginosa.

XX WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207272P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

DR N-PSDB; AAS54180.

XX New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -

XX Example 3; Seq ID No 11914; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 330 AA;

Query Match 1.5%; Score 8; DB 22; Length 330;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 EEEGALLA 60  
 |||||  
 Db 171 eeegalla 178

RESULT 13

ABB63962  
 ID ABB63962 standard; Protein; 542 AA.

AC ABB63962;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 18678.

XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL08065.

XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 genes from Drosophila and for elucidating cell signalling and cell-cell  
 interactions -

XX Disclosure; SEQ ID NO 18678; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
 capable of detecting 1000 or more genes from Drosophila. The invention is  
 useful in developmental biology and in elucidating cell signalling and  
 cell-cell interactions in higher eukaryotes for the development of  
 insecticides, therapeutics and pharmaceutical drugs. The invention  
 discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 sequences (ABL01840-ABL16175) and the encoded proteins  
 (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed  
 specification, but was obtained in electronic format directly from WIPO  
 at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 542 AA;

Query Match 1.5%; Score 8; DB 22; Length 542;

Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 PICTPSRS 127

|||||

Db 71 pictpsrs 78

RESULT 14

AAB51186

ID AAB51186 standard; Protein; 551 AA.

XX AAB51186;

XX 21-MAR-2001 (first entry)

XX Sea urchin sulfatase protein U SEQ ID NO:15.

XX Human; iduronate 2-sulfatase; IDS; glycosylated; gene therapy;  
 KW mucopolysaccharidosis inhibitor; IDS deficiency disorder;  
 KW Hunter syndrome; mucopolysaccharidosis type II.

XX Arbacia punctulata.

XX US6153188-A.

XX 28-NOV-2000.

XX 12-FEB-1999; 99US-0249003.

XX 17-DEC-1992; 92US-0991973.

XX 28-NOV-1994; 94US-0345212.

XX 12-NOV-1991; 91US-0790362.

XX (WOME-) WOMEN'S &amp; CHILDREN'S HOSPITAL.

XX Wilison PJ, Anson DS, Occhiodoro T, Bielicki J, Clements PR;

XX Hopwood JJ, Morris CP;

XX WPI; 2001-060076/07.

XX New highly glycosylated recombinant human iduronate 2-sulfatase (IDS)  
 useful for diagnosing or treating subjects suspected of having or  
 suffering from IDS deficiency disorders, e.g. Hunter syndrome  
 (mucopolysaccharidosis-II) -

XX Example 1; Fig 3; 53pp; English.

XX The present invention describes a recombinant human iduronate 2-sulfatase  
 (IDS). The recombinant IDS is more highly glycosylated than the naturally  
 occurring enzyme isolated from human tissue. The recombinant human IDS  
 can be produced in Chinese Hamster Ovary (CHO) cells or in a human cell.  
 The recombinant IDS comprises a fusion protein. It is a  
 mucopolysaccharidosis inhibitor and can be used in gene therapy.  
 The recombinant IDS is useful in treating and diagnosing subjects  
 suffering from or suspected of having IDS deficiency disorders, e.g.  
 Hunter syndrome (mucopolysaccharidosis type II). The present sequence  
 represents a sea urchin sulfatase which is used in comparison with the  
 IDS sequence in an example from the present invention.

XX Sequence 551 AA;

Query Match 1.5%; Score 8; DB 22; Length 551;

Best Local Similarity 100.0%; Pred. No. 41;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 173 MVGKWHLG 180

|||||

Db 153 mvgkwhlg 160

RESULT 15

ABB63246

ID ABB63246 standard; Protein; 579 AA.

XX ABB63246;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 16530.

XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX Drosophila melanogaster.  
OS WO200171042-A2.  
XX 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US09231.  
PF 23-MAR-2000; 2000US-191637P.  
XX 11-JUL-2000; 2000US-0614150.  
PR (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI WPI; 2001-656860/75.  
XX N-PSDB; ABL07349.  
DR New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
PT Disclosure; SEQ ID NO 16530; 21pp + Sequence Listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX Sequence 579 AA;  
SQ

Query Match 1.5%; Score 8; DB 22; Length 579;  
Best Local Similarity 100.0%; Pred. No. 43;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 174 VGKWHLGF 181  
Db 127 VGKWHLGF 134

Search completed: July 23, 2002, 19:08:41  
Job time: 347 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 15:52:14 ; Search time 356.68 Seconds  
(without alignments)  
10907.615 Million cell updates/sec

Title: US-09-495-823-8  
Perfect score: 2266  
Sequence: 1 caccgctcgccacgcgtc.....tgccacctgtgcccgaattc 2266

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1877	82.8	2253	AAH46863	Human 23553 sulfat
2	1549	68.4	1800	AAH46867	Human 23553 sulfat
3	1526	67.3	4639	AA545962	Human DNA encoding
4	1526	67.3	4640	AA545962	Human PRO708 (UNQ3
5	1526	67.3	4650	AA233987	Human PRO708 nucle
6	890	39.3	1375	AA522726	Human cDNA encodin
7	590	26.0	1340	AA522490	Human cDNA encodin
8	537	23.7	562	AAH99356	Human protein enco
9	479	21.1	729	AAA02344	Human colon cancer

10	434	19.2	629	22	AA541205	cDNA encoding nove
11	55	2.4	294	21	AAA00278	Human colon cancer
12	44	1.9	1082	21	AAA02370	Human colon cancer
13	32	1.4	53	20	AAZ33990	Human PRO708 hybri
14	32	1.4	53	21	AAZ78675	Human PRO708 hybri
15	26	1.1	774	22	AAK63748	Human immune/haema
16	26	1.1	2029	22	AAK52385	Human polynucleoti
17	25	1.1	2613	21	AAZ77614	Human cancer assoc
18	25	1.1	2648	22	AA534914	cDNA encoding nove
19	24	1.1	24	20	AAZ33988	Human PRO708 PCR f
20	24	1.1	24	20	AAZ33989	Human PRO708 PCR f
21	24	1.1	24	21	AAZ78673	Human PRO708 rever
22	24	1.1	24	21	AAZ78674	Human PRO708 rever
23	24	1.1	357	22	AAH36471	Human colon cancer
24	24	1.1	405	22	AAK57023	Human immune/haema
25	24	1.1	457	22	AAI83302	Human polynucleoti
26	24	1.1	1215	20	AAK00620	Human secreted pro
27	24	1.1	1316	21	AAK95796	Apoptosis related
28	24	1.1	1325	22	AAZ28772	Human immunoglobul
29	24	1.1	1403	22	AAZ05310	Human secreted pro
30	24	1.1	1450	22	ABA06520	Human cDNA SEQ ID
31	24	1.1	1451	22	ABA06709	Human cDNA SEQ ID
32	24	1.1	1451	22	AAZ28838	Human immunoglobul
33	24	1.1	1451	22	AAK61788	Human immune/haema
34	24	1.1	1578	22	AAZ33072	DNA encoding human
35	24	1.1	2167	22	AAI81669	Human ADAM protein
36	24	1.1	3133	22	AAF83398	P. chrysogenum ABC
37	24	1.1	23041	22	AAZ26779	Human genomic DNA
38	24	1.1	23135	22	AAZ26780	Human genomic DNA
39	24	1.1	23135	22	AAK81080	Human immune/haema
40	23	1.0	321	22	AAK56582	Human immune/haema
41	23	1.0	359	22	ABA08227	Human CDR62 homolo
42	23	1.0	360	22	AAI90981	Human polynucleoti
43	23	1.0	378	22	AAI92024	Human polynucleoti
44	23	1.0	395	22	AAI33263	Human polynucleoti
45	23	1.0	398	22	AAI85187	Human polynucleoti

ALIGNMENTS

RESULT 1

AAH46863

ID AAH46863 standard; cDNA; 2253 BP.

AC AAH46863;

XX

XX 25-SEP-2001 (first entry)

DT

XX Human 23553 sulfatase polypeptide encoding cDNA.

DE

XX Sulfatase; 23553 sulfatase; nootropic; neuroprotective; antibacterial;

KW antiinflammatory; vasotropic; antitumor; gene therapy; human; ss.

KW

XX Homo sapiens.

OS

XX	Key	Location/Qualifiers
XX	CDS	324..2123
XX	FT	/\*tag= a
XX	FT	/product= "23553 sulfatase"
XX	XX	WO200155411-A2.
XX	XX	02-AUG-2001.
XX	XX	31-JAN-2001; 2001WO-US03266.
XX	XX	31-JAN-2000; 2000US-0495823.
XX	XX	(MILL-) MILLENNIUM PHARM INC.
XX	XX	Glucksman MA, Williamson M, Rudolph-Owen LA, Tsai F;
XX	XX	



Qy 1741 gcaacctggaccgaacccggtggcacaatgaacgagatcacctcgtcaactggcaaaagt 1800  
|||||  
Db 1741 gcaacctggaccgaacccggtggcacaatgaacgagatcacctcgtcaactggcaaaagt 1800  
  
Qy 1801 tatggcttttcaacatacacccgaccatagagaggtgagactatcaacaggtatc 1860  
|||||  
Db 1801 tatggcttttcaacatacacccgaccatagagaggtgagactatcaacaggtatc 1860  
  
Qy 1861 caggaatcgtgaagaagctctacgaggtctctcacagtccaacaaaactcagtgccgg 1920  
|||||  
Db 1861 caggaatcgtgaagaagctctacgaggtctctcacagtccaacaaaactcagtgccgg 1920  
  
Qy 1921 tcaggtatcccccaaaagaccaccagaagtaaaccttagctcaatggaggggtctagggac 1980  
|||||  
Db 1921 tcaggtatcccccaaaagaccaccagaagtaaaccttagctcaatggaggggtctagggac 1980  
  
Qy 1981 catggtatagaggaagcaagcaaaaagcaagcaaaaatcagctgagaaaaagc 2040  
|||||  
Db 1981 catggtatagaggaagcaagcaaaaagcaagcaaaaatcagctgagaaaaagc 2040  
  
Qy 2041 aaaagaaagc-aaaaaagaagaagaacagcagagaagcaggtctcaggttcaacttgcc 2099  
|||||  
Db 2041 aaaagaaagc-aaaaaagaagaagaacagcagagaagcaggtctcaggttcaacttgcc 2100  
  
Qy 2100 attcaggtgttactgttgataagcacacaaatattctctgttgggttaaacctttaacagt 2159  
|||||  
Db 2101 attcaggtgttactgttgataagcacacaaatattctctgttgggttaaacctttaacagt 2160  
  
Qy 2160 tcttatcttcatctgttcttagtaaacacagcaaaattggctcgataatcgtggc 2219  
|||||  
Db 2161 tcttatcttcatctgttcttagtaaacacagcaaaattggctcgataatcgtggc 2220  
  
Qy 2220 ctaagcgtcaggtgttctcatgctgtgccac 2252  
|||||  
Db 2221 ctaagcgtcaggtgttctcatgctgtgccac 2253

RESULT 2  
AAH46867

ID AAH46867 standard; cDNA; 1800 BP.

XX AAH46867;

XX 25-SEP-2001 (first entry)

XX Human 23553 sulfatase polypeptide coding sequence.

XX Sulfatase; 23553 sulfatase; nootropic; neuroprotective; antibacterial;  
KW antiinflammatory; vasotropic; antitumor; gene therapy; human; ss.

XX Homo sapiens.

XX Key Location/Qualifiers  
FH 1..1800  
FT CDS /\*tag= a

FT /\*product= "23553 sulfatase"

XX WO200155411-A2.

XX 02-AUG-2001.

XX 31-JAN-2001; 2001WO-US03266.

XX 31-JAN-2000; 2000US-0495823.

XX (MILL-) MILLENNIUM PHARM INC.

XX Glucksmann MA, Williamson M, Rudolph-Owen LA, Tsai F;

XX WPI; 2001-476214/51.

XX P-PSDB; AAB85483.

Qy 984 gactgtatgaaacgacaatgctgctgggactatgacaatgggataactccacag 1043

PT Novel human sulfatase polypeptides useful for treating and diagnosing  
PT sulfatase-related disorders such as cerebrovascular diseases, acute  
PT meningitis, multiple sclerosis, degenerative diseases and tumor -  
XX Claim 2; Fig 15; 180pp; English.  
XX  
CC The invention provides 22438, 23553, 25278 or 26212 human sulfatase  
CC polypeptides and polynucleotides. The sulfatase genes and polypeptides  
CC are useful for treating disorders involving the brain such as  
CC cerebrovascular diseases, infections such as acute meningitis,  
CC demyelinating diseases including multiple sclerosis, degenerative  
CC diseases affecting the cerebral cortex including Alzheimer's disease  
CC and Pick disease, spinocerebellar degenerations including spinocerebellar  
CC ataxias including Friedreich ataxia, and ataxia telangiectasia,  
CC degenerative diseases affecting motor neurons including amyotrophic  
CC lateral sclerosis, inborn errors of metabolism such as leukodystrophies,  
CC toxic and acquired metabolic diseases, including vitamin deficiencies,  
CC and neurocutaneous syndromes (phakamatoses) including neurofibromatosis.  
CC The present sequence represents a human 23553 sulfatase polypeptide  
CC coding sequence.  
XX  
SQ Sequence 1800 BP; 534 A; 434 C; 423 G; 409 T; 0 other;

Query Match 68.4%; Score 1549; DB 22; Length 1800;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1649; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 324 atggctccagggtgtgtcggggcatccgctcgcgtcttccacaggcctgtgtctgt 383  
|||||

Db 1 atggctccagggtgtgtcggggcatccgctcgcgtcttccacaggcctgtgtctgt 60  
|||||

Qy 384 cctgaaagatgactagcaatggggtgctgagagattctgctctgctcctcact 443  
|||||

Db 61 cctgaaagatgactagcaatggggtgctgagagattctgctctgctcctcact 120  
|||||

Qy 444 tatgttacctgtcctggggcaggccttagaagaggaggaaggagggccttactagct 503  
|||||

Db 121 tatgttacctgtcctggggcaggccttagaagaggaggaaggagggccttactagct 180  
|||||

Qy 504 caagctgagagaaactagagccacgaactccacctccacagcccatctcatttcc 563  
|||||

Db 181 caagctgagagaaactagagccacgaactccacctccacagcccatctcatttcc 240  
|||||

Qy 564 atcctagcggatgacagggtatttagagatgtgggttaccacggatctgagattaaaca 623  
|||||

Db 241 atcctagcggatgacagggtatttagagatgtgggttaccacggatctgagattaaaca 300  
|||||

Qy 624 cctactcttgacaagctcgtcggcgaaggagttaaactggagaactactatgtccagct 683  
|||||

Db 301 cctactcttgacaagctcgtcggcgaaggagttaaactggagaactactatgtccagct 360  
|||||

Qy 684 atttgacacatccaggagtcagtttattactggaagatcagatacacacccggactt 743  
|||||

Db 361 atttgacacatccaggagtcagtttattactggaagatcagatacacacccggactt 420  
|||||

Qy 744 caacattctatcataagacctaccacccaaactgtttacctcttgacaaatgccacctta 803  
|||||

Db 421 caacattctatcataagacctaccacccaaactgtttacctcttgacaaatgccacctta 480  
|||||

Qy 804 cctcagaactgaaggaggttgattatcaacgcataatggtcggaaaaatggcacttgggt 863  
|||||

Db 481 cctcagaactgaaggaggttgattatcaacgcataatggtcggaaaaatggcacttgggt 540  
|||||

Qy 864 ttttcagaaaaaatgcatgccccccagagaagatttgatcaccttttttggttccctt 923  
|||||

Db 541 ttttcagaaaaaatgcatgccccccagagaagatttgatcaccttttttggttccctt 600  
|||||

Qy 924 ttgggaagtgggtattactatcacactacaatgtgacagtcctggagatgtgtgctat 983  
|||||

Db 601 ttgggaagtgggtattactatcacactacaatgtgacagtcctggagatgtgtgctat 660  
|||||

Db 661 gactgtatgaagacaatgctgctgggactatgacaatggcatatactccacag 720  
QY 1044 atgtacactcagagtagcacgaaatcttagcttcccaataaccacaaagcctatatt 1103  
Db 721 atgtacactcagagtagcacgaaatcttagcttcccaataaccacaaagcctatatt 780  
QY 1104 ttatatattgctatcaagctgttcatccaccactgcaagctcctcggaaggtattcgaa 1163  
Db 781 ttatatattgctatcaagctgttcatccaccactgcaagctcctcggaaggtattcgaa 840  
QY 1164 cactaccgattccattatcaacataaacaggaggagatgctgcccattgcttccctgctta 1223  
Db 841 cactaccgattccattatcaacataaacaggaggagatgctgcccattgcttccctgctta 900  
QY 1224 gatgaagcaatcaacaacgtgacattggctctaaagacttatggtttctataacaacagc 1283  
Db 901 gatgaagcaatcaacaacgtgacattggctctaaagacttatggtttctataacaacagc 960  
QY 1284 attatcatttactcttcagataatggtggccagctacggcagggagtaactggcct 1343  
Db 961 attatcatttactcttcagataatggtggccagctacggcagggagtaactggcct 1020  
QY 1344 ctacaggttagcaaggacaattatgggaaggaggtatccggctgtaggctttgtgcat 1403  
Db 1021 ctacaggttagcaaggacaattatgggaaggaggtatccggctgtaggctttgtgcat 1080  
QY 1404 agcccaactctgaagaaaggaacagtggttaaggaacctgtgcacataactgactgg 1463  
Db 1081 agcccaactctgaagaaaggaacagtggttaaggaacctgtgcacataactgactgg 1140  
QY 1464 taccacactctatttcaactgctgaaggacagattgatgaggacattcaactagatgac 1523  
Db 1141 taccacactctatttcaactgctgaaggacagattgatgaggacattcaactagatgac 1200  
QY 1524 tatgatattgggaagcaataagttaggggtcttcgctcaccgcagtagatatttgcac 1583  
Db 1201 tatgatattgggaagcaataagttaggggtcttcgctcaccgcagtagatatttgcac 1260  
QY 1584 aaattgaccccatatcaccaagcaaaaaatgctcctgggcagcaggtatggatc 1643  
Db 1261 aaattgaccccatatcaccaagcaaaaaatgctcctgggcagcaggtatggatc 1320  
QY 1644 tggaaactgcgaactcagtcagccatcagagtgagtcagcactggaaattgcttacaggaat 1703  
Db 1321 tggaaactgcgaactcagtcagccatcagagtgagtcagcactggaaattgcttacaggaat 1380  
QY 1704 cctggtacagcagactgggtcccccctcagttcttcagcaaacctgggaccgaaccggtgg 1763  
Db 1381 cctggtacagcagactgggtcccccctcagttcttcagcaaacctgggaccgaaccggtgg 1440  
QY 1764 cacaatgaacggtacactcgtcaactggcaaaagtgtatgcttttcaacatacagcc 1823  
Db 1441 cacaatgaacggtacactcgtcaactggcaaaagtgtatgcttttcaacatacagcc 1500  
QY 1824 gaccatattgaggggtgacacctatctaacaggtatccaggaatcgtgaagagctccta 1883  
Db 1501 gaccatattgaggggtgacacctatctaacaggtatccaggaatcgtgaagagctccta 1560  
QY 1884 cggaggtctctacagttcaacaaaaactgcagtgccgggtcaggtatccccccaaagacccc 1943  
Db 1561 cggaggtctctacagttcaacaaaaactgcagtgccgggtcaggtatccccccaaagacccc 1620  
QY 1944 aqaagtaaccttagctcaatggaggggtct 1974  
Db 1621 aqaagtaaccttagctcaatggaggggtct 1651

RESULT 3  
AAS45962  
ID: AAS45962 standard; cDNA; 4639 BP.  
XX  
AC AAS45962;

XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Human DNA encoding PRO polypeptide sequence #38.  
XX  
KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; ss;  
KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;  
KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;  
KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder;  
KW PCR primer.  
XX  
OS Homo sapiens.  
XX  
PN WO200168848-A2.  
XX  
PD 20-SEP-2001.  
XX  
PF 28-FEB-2001; 2001WO-US06520.  
XX  
PR 01-MAR-2000; 2000WO-US05601.  
PR 02-MAR-2000; 2000WO-US05841.  
PR 03-MAR-2000; 2000US-187202P.  
PR 06-MAR-2000; 2000US-186968P.  
PR 14-MAR-2000; 2000US-189320P.  
PR 14-MAR-2000; 2000US-189328P.  
PR 15-MAR-2000; 2000WO-US06884.  
PR 21-MAR-2000; 2000US-190828P.  
PR 21-MAR-2000; 2000US-191007P.  
PR 21-MAR-2000; 2000US-191048P.  
PR 21-MAR-2000; 2000US-191314P.  
PR 28-MAR-2000; 2000US-192655P.  
PR 29-MAR-2000; 2000US-193032P.  
PR 30-MAR-2000; 2000US-193053P.  
PR 30-MAR-2000; 2000WO-US08439.  
PR 04-APR-2000; 2000US-194449P.  
PR 04-APR-2000; 2000US-194647P.  
PR 11-APR-2000; 2000US-195975P.  
PR 11-APR-2000; 2000US-196000P.  
PR 11-APR-2000; 2000US-196187P.  
PR 11-APR-2000; 2000US-196690P.  
PR 11-APR-2000; 2000US-196820P.  
PR 18-APR-2000; 2000US-198121P.  
PR 18-APR-2000; 2000US-198585P.  
PR 25-APR-2000; 2000US-199397P.  
PR 25-APR-2000; 2000US-199550P.  
PR 25-APR-2000; 2000US-199654P.  
PR 03-MAY-2000; 2000US-201516P.  
PR 17-MAY-2000; 2000WO-US13705.  
PR 22-MAY-2000; 2000WO-US14042.  
PR 30-MAY-2000; 2000WO-US14941.  
PR 02-JUN-2000; 2000WO-US15264.  
PR 05-JUN-2000; 2000US-209832P.  
PR 28-JUL-2000; 2000WO-US20710.  
PR 22-AUG-2000; 2000US-0644848.  
PR 24-AUG-2000; 2000WO-US23328.  
PR 08-NOV-2000; 2000WO-US30952.  
PR 01-DEC-2000; 2000WO-US32678.  
PR 20-DEC-2000; 2000WO-US34956.  
XX  
(GETH ) GENENTECH INC.  
XX  
PI Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;  
PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;  
XX  
XX WPI: 2001-602746/68.  
DR P-PSDB; AAU29061.  
XX  
PT Novel nucleic acids encoding PRO polypeptides, used to diagnose the  
PT presence of tumours, such as prostate and breast tumours, in mammals and  
XX to screen for modulators of the compounds -  
XX  
PS Claim 2; Fig 75; 774pp; English.  
XX





Qy 922 ttttgggaagtgggattactatatacacactacaaaatgtgacagtcctctggatgtgtgct 981  
Db 1489 ttttgggaagtgggattactatatacacactacaaaatgtgacagtcctctggatgtgtgct 1548  
Qy 982 atgacttgtatgaaaacagacaatgctgctgggactatgacaatggcatatactccacac 1041  
Db 1549 atgacttgtatgaaaacagacaatgctgctgggactatgacaatggcatatactccacac 1608  
Qy 1042 agatgacactcagagagatcacagaaatcttagcttccataaaccacaaagcctatat 1101  
Db 1609 agatgacactcagagagatcacagaaatcttagcttccataaaccacaaagcctatat 1668  
Qy 1102 ttttatattgcttatcaagctgttccattccaccctgcaagctcctggcagggtatttcg 1161  
Db 1669 ttttatattgcttatcaagctgttccattccaccctgcaagctcctggcagggtatttcg 1728  
Qy 1162 aacactaccgatccattatacaataaagagagagatatgctgccatgctttctctgct 1221  
Db 1729 aacactaccgatccattatacaataaagagagagatatgctgccatgctttctctgct 1788  
Qy 1222 tagatgaagcaatacaaacctgacattgctctaaagacttatggtttctataacaaca 1281  
Db 1789 tagatgaagcaatacaaacctgacattgctctaaagacttatggtttctataacaaca 1848  
Qy 1282 gcattatcatttactcttcagataatggtgcccagctcacggcaggagtgtaactggc 1341  
Db 1849 gcattatcatttactcttcagataatggtgcccagctcacggcaggagtgtaactggc 1908  
Qy 1342 ctctcagaggtagcaaaaggaacatatattgggaagaggagatccggctgtaggcttttgc 1401  
Db 1909 ctctcagaggtagcaaaaggaacatatattgggaagaggagatccggctgtaggcttttgc 1968  
Qy 1402 atagcccactctgaaaaaaacaggaagacagtgtgtaaggaacctgtgcacatcacctgact 1461  
Db 1969 atagcccactctgaaaaaaacaggaagacagtgtgtaaggaacctgtgtcacatcacctgact 2028  
Qy 1462 ggtaccccaactctcatttcaactggtgctgaagacagattgatgagacatttcaactagatg 1521  
Db 2029 ggtaccccaactctcatttcaactggtgctgaagacagattgatgagacatttcaactagatg 2088  
Qy 1522 gctatgatattgggagacataagtagaggtcttctgcctcaccocagtagatatatttgc 1581  
Db 2089 gctatgatattgggagacataagtagaggtcttctgcctcaccocagtagatatatttgc 2148  
Qy 1582 ataacattgaccccc-tataaccaaggcaaaaatggtctctggcagcaggtatggga 1641  
Db 2149 ataacattgaccccc-tataaccaaggcaaaaatggtctctggcagcaggtatggga 2207  
Qy 1642 tctggaacactgcaatccagtcagccatcagagtcagactggaattgtttacaggaa 1701  
Db 2208 tctggaacactgcaatccagtcagccatcagagtcagactggaattgtttacaggaa 2267  
Qy 1702 atcctggtcacagcagactgggtcccccctcagttcttcaagaacctgggacccggaaccggt 1761  
Db 2268 atcctggtcacagcagactgggtcccccctcagttcttcaagaacctgggacccggaaccggt 2327  
Qy 1762 ggcacaaatgaacgcatcaccctcgtcaactggcaaaaagtgtatggcttttcaacatcacag 1821  
Db 2328 ggcacaaatgaacgcatcacccttgtcaactggcaaaaagtgtatggcttttcaacatcacag 2387  
Qy 1822 ccgaccccatatgagaggtggacctatatacaggtatccagggaatcgtgaagaagctcc 1881  
Db 2388 ccgaccccatatgagaggtggacctatatacaggtatccagggaatcgtgaagaagctcc 2447  
Qy 1882 tacggaggctctcagttcaacaaaactcagtcgcggtcaggtatccccccaaagacc 1941  
Db 2448 tacggaggctctcagttcaacaaaactcagtcgcggtcaggtatccccccaaagacc 2507  
Qy 1942 ccagaagttaacccttaggctcaatgaggggtct 1974  
Db 2508 ccagaagttaacccttaggctcaatgaggggtct 2540

RESULT 5  
AAZ33987  
ID AAZ33987 standard; cDNA; 4650 BP.  
XX AAZ33987;  
XX 07-DEC-1999 (first entry)  
XX Human PRO708 nucleotide sequence.  
XX Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;  
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;  
KW secreted protein; transmembrane protein; ss.  
XX Homo sapiens.  
XX WO9946281-A2.  
XX 16-SEP-1999.  
XX 08-MAR-1999; 99WO-US05028.  
XX 10-MAR-1998; 98US-0077450.  
PR 11-MAR-1998; 98US-0077632.  
PR 11-MAR-1998; 98US-0077641.  
PR 11-MAR-1998; 98US-0077649.  
PR 12-MAR-1998; 98US-0077791.  
PR 13-MAR-1998; 98US-0078004.  
PR 17-MAR-1998; 98US-0040220.  
PR 20-MAR-1998; 98US-0078886.  
PR 20-MAR-1998; 98US-0078910.  
PR 20-MAR-1998; 98US-0078936.  
PR 20-MAR-1998; 98US-0078939.  
PR 25-MAR-1998; 98US-0079294.  
PR 26-MAR-1998; 98US-0079656.  
PR 27-MAR-1998; 98US-0079663.  
PR 27-MAR-1998; 98US-0079664.  
PR 27-MAR-1998; 98US-0079728.  
PR 27-MAR-1998; 98US-0079786.  
PR 30-MAR-1998; 98US-0079920.  
PR 30-MAR-1998; 98US-0079923.  
PR 31-MAR-1998; 98US-0080105.  
PR 31-MAR-1998; 98US-0080107.  
PR 31-MAR-1998; 98US-0080165.  
PR 31-MAR-1998; 98US-0080194.  
PR 01-APR-1998; 98US-0080327.  
PR 01-APR-1998; 98US-0080328.  
PR 01-APR-1998; 98US-0080333.  
PR 01-APR-1998; 98US-0080334.  
PR 08-APR-1998; 98US-0081049.  
PR 08-APR-1998; 98US-0081070.  
PR 08-APR-1998; 98US-0081071.  
PR 09-APR-1998; 98US-0081195.  
PR 09-APR-1998; 98US-0081203.  
PR 09-APR-1998; 98US-0081229.  
PR 15-APR-1998; 98US-0081817.  
PR 15-APR-1998; 98US-0081838.  
PR 15-APR-1998; 98US-0081952.  
PR 15-APR-1998; 98US-0081955.  
PR 21-APR-1998; 98US-0082568.  
PR 21-APR-1998; 98US-0082569.  
PR 22-APR-1998; 98US-0082700.  
PR 22-APR-1998; 98US-0082704.  
PR 22-APR-1998; 98US-0082804.  
PR 23-APR-1998; 98US-0082767.  
PR 23-APR-1998; 98US-0082796.  
PR 27-APR-1998; 98US-0083336.  
PR 28-APR-1998; 98US-0083322.  
PR 29-APR-1998; 98US-0083392.  
PR 29-APR-1998; 98US-0083495.  
PR 29-APR-1998; 98US-0083496.







QY 1162 aactaccgattccattatcaacataaacacagagagagatgctgcttctgct 1221  
DB 1729 aactaccgattccattatcaacataaacacagagagatgctgcttctgct 1788  
QY 1222 tagatgaagcaatcaacacagtgacattggtcttaaaagacttatggtttctataacaaca 1281  
DB 1789 tagatgaagcaatcaacacagtgacattggtcttaaaagacttatggtttctataacaaca 1848  
QY 1282 gattattacttactcttcagataatggtgccagctacgagcagagggagtaactggc 1341  
DB 1849 gattattacttactcttcagataatggtgccagctacgagcagagggagtaactggc 1908  
QY 1342 ctctcagaggtagcaaaagaaacatatgttggaagggaggtccggctgtaggctttgtgc 1401  
DB 1909 ctctcagaggtagcaaaagaaacatatgttggaagggaggtccggctgtaggctttgtgc 1968  
QY 1402 atagcccacttctgaaaaaacaggaaacagtggttaagaaacctgtgcacatcactgact 1461  
DB 1969 atagcccacttctgaaaaaacaggaaacagtggttaagaaacctgtgcacatcactgact 2028  
QY 1462 ggtaccccactctctattcactggtcgaagacagattgatgagacattcaactagatg 1521  
DB 2029 ggtaccccactctctattcactggtcgaagacagattgatgagacattcaactagatg 2088  
QY 1522 gctatgatctggagaccataagtggaggtcttcgctcaccgccagtagatatgttgc 1581  
DB 2089 gctatgatctggagaccataagtggaggtcttcgctcaccgccagtagatatgttgc 2148  
QY 1582 ataacattgacccc-tatacaccaggcaaaaaatggctcctgggcagcaggctatggga 1641  
DB 2149 ataacattgacccc-tatacaccaggcaaaaaatggctcctgggcagcaggctatggga 2207  
QY 1642 tctgaaacactgcaatccagtcagccatcagagtcagactggaattgcttacagaa 1701  
DB 2208 tctgaaacactgcaatccagtcagccatcagagtcagactggaattgcttacagaa 2267  
QY 1702 atcttggtacagcagactgggtcccccctcagttcttccagaaacctgggaccgaccggt 1761  
DB 2268 atcttggtacagcagactgggtcccccctcagttcttccagaaacctgggaccgaccggt 2327  
QY 1762 ggcacaaatgaacggatccctgtaacctgcaaaagtgtatggttttcaacatcacag 1821  
DB 2328 ggcacaaatgaacggatccctgtaacctgcaaaagtgtatggttttcaacatcacag 2387  
QY 1822 ccgacccatatgagaggtgaccttatcaacaggtatccaggaatcgtgaagaagctcc 1881  
DB 2388 ccgacccatatgagaggtgaccttatcaacaggtatccaggaatcgtgaagaagctcc 2447  
QY 1882 tacggaggctctcacagttcaacaaaactgcagtcgggtcaggtatccccccaaagacc 1941  
DB 2448 tacggaggctctcacagttcaacaaaactgcagtcgggtcaggtatccccccaaagacc 2507  
QY 1942 ccagaagtacccctagggtcaatgaggggtct 1974  
DB 2508 ccagaagtacccctagggtcaatgaggggtct 2540

RESULT 6  
ID AAS22726 standard; cDNA; 1375 BP.  
XX  
AC AAS22726;  
XX  
DT 24-OCT-2001 (first entry)  
XX  
DE Human cDNA encoding a novel human protein #292.  
XX  
KW Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;  
KW immunomodulatory; cytotatic; neuroprotective; vulnerary; nootropic;  
KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;  
KW antibacterial; anti allergic; dermatological; haemostatic; antiasthmatic;  
KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;

KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;  
KW tissue regeneration; immune disorder.  
OS Homo sapiens.  
XX WO200155437-A2.  
XX PD 02-AUG-2001.  
XX 25-JAN-2001; 2001WO-US02623.  
XX 25-JAN-2000; 2000US-0491404.  
XX (HYSE-) HYSEQ INC.  
PA Tang YT, Liu C, Drmanac RT;  
PI WPI; 2001-451939/48.  
DR P-PSDB; AAU14421.  
XX Isolated polypeptides useful for treating anti-inflammatory diseases,  
PT nervous system disorders, and for regenerating bone and cartilage -  
XX Claim 1; Page 699; 894pp; English.  
XX The invention relates to polynucleotides encoding novel human  
CC proteins or their active domains. The polypeptides, polynucleotides and  
CC antibodies raised against the polypeptides are used in a method of  
CC treatment of a mammal and prevention of disorders caused by the aberrant  
CC protein expression or activity. The polypeptides can be used as  
CC molecular weight markers, food supplements, and in antibody production.  
CC The polypeptides are used to identify compounds which bind to the  
CC polypeptides. Polynucleotides of the invention are used as probes and  
CC primers for sequencing, for chromosome or gene mapping, in the  
CC production of recombinant proteins, and in generating anti-sense DNA or  
CC RNA and in gene therapy. Polypeptides of the invention can be used to  
CC target drugs to a tumour, in assays to determine biological activity, to  
CC raise antibodies/elicit an immune response, to determine quantitative  
CC protein levels, as tissue markers, and to isolate receptors or ligands.  
CC Polypeptides of the invention may also be useful in treating platelet  
CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,  
CC ligament and/or nerve tissue, wound healing, treating burns, promoting  
CC the proliferation, differentiation and survival of stem cells, as a  
CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,  
CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral  
CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or  
CC fungal infection or from autoimmunity, cancer, allergy, asthma,  
CC graft-versus-host disease, eczema, haemophilia, thrombosis,  
CC anti-inflammatory diseases, nervous system disorders, and infection.  
XX The present sequence encodes a protein of the invention.  
SQ Sequence 1375 BP; 399 A; 336 C; 360 G; 280 T; 0 other;  
  
Query Match 39.3%; Score 890; DB 22; Length 1375;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 890; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 369 caggcctgtgtctcctggaaaagatgctagcaatggggcgctggcagattctggatc 428  
DB 484 caggcctgtgtctcctggaaaagatgctagcaatggggcgctggcagattctggatc 543  
QY 429 ctctgcctctcacttatgtttacctgtctctggggcaggccttagaagaggaagaa 488  
DB 544 ctctgcctctcacttatgtttacctgtctctggggcaggccttagaagaggaagaa 603  
QY 489 ggggccttactagctcaagctggagaaactagagccagcacaaacttcccactccccag 548  
DB 604 ggggccttactagctcaagctggagaaactagagccagcacaaacttcccactccccag 663  
QY 549 cccatctcatttcatctcctcagcgatgatcaggggatttagagatgtgggttaccacgga 608  
DB 664 cccatctcatttcatctcctcagcgatgatcaggggatttagagatgtgggttaccacgga 723

```
QY 609 tctgagattaaacacccactactcttgacaagctcgtgcggaaggaggttaaaactggagac 668
DB 724 tctgagattaaacacccactactcttgacaagctcgtgcggaaggaggttaaaactggagac 783
QY 669 tactatgtccagcctatttgacacaccatccaggagtcagttatttactcggaaagtatcag 728
DB 784 tactatgtccagcctatttgacacaccatccaggagtcagttatttactcggaaagtatcag 843
QY 729 atacacacggagctcaacattctatcataagacacctacccaacccaactgtttactcttg 788
DB 844 atacacacggagctcaacattctatcataagacacctacccaacccaactgtttactcttg 903
QY 789 gacaatgcacccctaccctcagaacactgaaggaggttgatattcaacacatattgctcga 848
DB 904 gacaatgcacccctaccctcagaacactgaaggaggttgatattcaacacatattgctcga 963
QY 849 aaatggcaacttgggtttttacagaaaaaagtcatgccacacagaagaggtttgatacc 908
DB 964 aaatggcaacttgggtttttacagaaaaaagtcatgccacacagaagaggtttgatacc 1023
QY 909 tttttgttcccttttgggaagtgggattactactatcacactacaaatgtgacagtctc 968
DB 1024 tttttgttcccttttgggaagtgggattactactatcacactacaaatgtgacagtctc 1083
QY 969 gggatgtggctgactgttatgaaaacgacaaatgctgctgggactatgacaaatggc 1028
DB 1084 gggatgtggctgactgttatgaaaacgacaaatgctgctgggactatgacaaatggc 1143
QY 1029 atatactccacacagatgtacactcagagagtcagcaaatatttgcctccataacccc 1088
DB 1144 atatactccacacagatgtacactcagagagtcagcaaatatttgcctccataacccc 1203
QY 1089 acaagacctatttttatatttcctcatcaagctgttcattccaccactgcaagtctc 1148
DB 1204 acaagacctatttttatatttcctcatcaagctgttcattccaccactgcaagtctc 1263
QY 1149 ggcaggtatttcgaactaccgattccattatcaacataaaacagagagatgtctgcc 1208
DB 1264 ggcaggtatttcgaactaccgattccattatcaacataaaacagagagatgtctgcc 1323
QY 1209 atgctttcctgcttgatgaagcaatcaacacagtcgacattggctctaaa 1258
DB 1324 atgctttcctgcttgatgaagcaatcaacacagtcgacattggctctaaa 1373

RESULT 7
AAS22490
ID AAS22490 standard; cDNA; 1340 BP.
AC
XX
AC AAS22490;
XX
DB 24-OCT-2001 (first entry)
XX
DE Human cDNA encoding a novel human protein #56.
XX
KW Human; novel protein; ss; Antianaemic; osteopathic; antiinflammatory;
KW immunomodulatory; cytostatic; neuroprotective; vulnery; nootropic;
KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW tissue regeneration; immune disorder.
XX
OS Homo sapiens.
XX
PN WO200155437-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02623.
XX
PR 25-JAN-2000; 2000US-0491404.
```

```
XX (HYSE-) HYSEQ INC.
PA Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-451939/48.
DR P-PSDB; AAU14185.
XX Isolated polypeptides useful for treating anti-inflammatory diseases,
XX nervous system disorders, and for regenerating bone and cartilage -
PS Claim 1; Page 256-257; 894pp; English.
XX The invention relates to polynucleotides encoding novel human
XX proteins or their active domains. The polypeptides, polynucleotides and
XX antibodies raised against the polypeptides are used in a method of
XX treatment of a mammal and prevention of disorders caused by the aberrant
XX protein expression or activity. The polypeptides can be used as
XX molecular weight markers, food supplements, and in antibody production.
XX The polypeptides are used to identify compounds which bind to the
XX polypeptides. Polynucleotides of the invention are used as probes and
XX primers, for sequencing, for chromosome or gene mapping, in the
XX production of recombinant proteins, and in generating anti-sense DNA or
XX RNA and in gene therapy. Polypeptides of the invention can be used to
XX target drugs to a tumour, in assays to determine biological activity, to
XX raise antibodies/elicit an immune response, to determine quantitative
XX protein levels, as tissue markers, and to isolate receptors or ligands.
XX Polypeptides of the invention may also be useful in treating platelet
XX disorders, stem cell disorders, regenerating bone, cartilage, tendon,
XX ligament and/or nerve tissue, wound healing, treating burns, promoting
XX the proliferation, differentiation and survival of stem cells, as a
XX contraceptive, treating osteoporosis and osteoarthritis, anaemia,
XX Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral
XX sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
XX fungal infection or from autoimmunity, cancer, allergy, asthma,
XX graft-versus-host disease, eczema, haemophilia, thrombosis,
XX anti-inflammatory diseases, nervous system disorders, and infection.
XX The present sequence encodes a protein of the invention.
SQ Sequence 1340 BP; 374 A; 338 C; 369 G; 258 T; 1 other;

Query Match 26.0%; Score 590; DB 22; Length 1340;
Best Local Similarity 100.0%; Pred. No. 2.6e-272;
Matches 590; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 369 caggcctgtgtcctcctggaaagatgtagcaatggggcgctggcaggattctggatc 428
DB 484 caggcctgtgtcctcctggaaagatgtagcaatggggcgctggcaggattctggatc 543
QY 429 ctctgcctcctcaactattgttacctgtcctggggccaggccttagaagaggagaaagaa 488
DB 544 ctctgcctcctcaactattgttacctgtcctggggccaggccttagaagaggagaaagaa 603
QY 489 ggggccttactagctcaagctggagagaaactagaccagcacaaactccacctccacag 548
DB 604 ggggccttactagctcaagctggagagaaactagaccagcacaaactccacctccacag 663
QY 549 ccccatctcatttccatcctcctcagcagatcagcaggatttagagatgtgggttacacagga 608
DB 664 ccccatctcatttccatcctcctcagcagcaggatttagagatgtgggttacacagga 723
QY 609 tctgagattaaaaaacctactcttgcacagctcgtgcggaaggaggttaaaactggagaa 668
DB 724 tctgagattaaaaaacctactcttgcacagctcgtgcggaaggaggttaaaactggagaa 783
QY 669 tactatgtccagcctatttgacacaccatccaggagtcagttatttactcggaaagtatcag 728
DB 784 tactatgtccagcctatttgacacaccatccaggagtcagttatttactcggaaagtatcag 843
QY 729 atacacacggagctcaacattctatcataagacacctacccaacccaactgtttactcttg 788
DB 844 atacacacggagctcaacattctatcataagacacctacccaacccaactgtttactcttg 903
```

Qy 789 gacaatgccaccctacacgaaactgaagaggttggtatattcaacgcataatggtcgga 848  
|||||  
Db 904 gacaatgccaccctacacgaaactgaagaggttggtatattcaacgcataatggtcgga 963  
Qy 849 aaatggcaactgggtttttacagaaaagaatgcagtcgccaccagagaagaggttgatacc 908  
|||||  
Db 964 aaatggcaactgggtttttacagaaaagaatgcagtcgccaccagagaagaggttgatacc 1023  
Qy 909 tttttggtcccttttgggaagtgggattactatatacacactacaagt 958  
|||||  
Db 1024 tttttggtcccttttgggaagtgggattactatatacacactacaagt 1073

RESULT 8  
AAH99356  
ID AAH99356 standard; cDNA; 562 BP.  
XX  
AC AAH99356;  
XX  
DT 16-OCT-2001 (first entry)  
XX  
DE Human protein encoding cDNA sequence SEQ ID NO:191.  
XX  
KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
KW antibacterial; endocrine; cardiac; central nervous system; virucide;  
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
KW antiaggregant; haemostatic; vulnery; antiulcer; osteopathic; eczema;  
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;  
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
KW neurological disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200153455-A2.  
XX  
PD 26-JUL-2001.  
XX  
PF 22-DEC-2000; 2000WO-US35017.  
XX  
PR 23-DEC-1999; 99US-0471275.  
PR 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Tang YT, Liu C, Drmanac RT;  
XX  
DR WPI; 2001-457603/49.  
DR P-PSDB; AAM25415.  
XX  
PT Isolated human polynucleotides encoding polypeptides, useful for the  
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -  
XX  
PS Claim 1; Page 383; 1217pp; English.  
XX  
CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to  
CC AAM25963. The proteins can have activities based on the tissues and  
CC cells they are expressed in, such as: antiinflammatory; antirheumatic;  
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;  
CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;  
CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery;  
CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;  
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;  
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides

CC encoding them can be used in gene therapy, antisense therapy and vaccine  
CC production. The proteins and polynucleotides are useful for screening for  
CC agonists or antagonists of a protein and for the treatment and diagnosis  
CC of disorders associated with the activity of a protein e.g. inflammation,  
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,  
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,  
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,  
CC osteoporosis, severe combined immunodeficiency, eczema, allergic  
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and  
CC neurological disorders.  
XX

SQ Sequence 562 BP; 170 A; 132 C; 105 G; 155 T; 0 other;

Query Match 23.7%; Score 537; DB 22; Length 562;  
Best Local Similarity 100.0%; Pred. NO. 6.4e-247;  
Matches 537; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 722 gtatcagatacacaccggacttcaacattctatataagacacctcccaacccaactgttt 781  
|||||  
Db 24 gtatcagatacacaccggacttcaacattctatataagacacctcccaacccaactgttt 83

Qy 782 acctgtgacaatgccaccctacctcagaaactgaaggaggttggtatattcaacgcata 841  
|||||  
Db 84 acctgtgacaatgccaccctacctcagaaactgaaggaggttggtatattcaacgcata 143

Qy 842 ggtcggaaaaatggcaacttgggtttttacagaaaagaatgcagtcgccaccagagaagatt 901  
|||||  
Db 144 ggtcggaaaaatggcaacttgggtttttacagaaaagaatgcagtcgccaccagagaagatt 203

Qy 902 tgatacctttttgttcccttttgggaagtggggtattactatatacacactacaaatgtga 961  
|||||  
Db 204 tgatacctttttgttcccttttgggaagtggggtattactatatacacactacaaatgtga 263

Qy 962 cagtcctgggagtgtggtgctatgacttgtatgaaaaacgacaatgctgctgggactatga 1021  
|||||  
Db 264 cagtcctgggagtgtggtgctatgacttgtatgaaaaacgacaatgctgctgggactatga 323

Qy 1022 caatggcatatactccacacagatgtacactcagagatcacagaaatcttagcttccca 1081  
|||||  
Db 324 caatggcatatactccacacagatgtacactcagagatcacagaaatcttagcttccca 383

Qy 1082 taaccccaaaagcctatatattttatatattgctctataagctgttctaccactgca 1141  
|||||  
Db 384 taaccccaaaagcctatatattttatatattgctctataagctgttctaccactgca 443

Qy 1142 agctcctggcagggtatttcgaacactaccgattccattatcaacataaagaggagata 1201  
|||||  
Db 444 agctcctggcagggtatttcgaacactaccgattccattatcaacataaagaggagata 503

Qy 1202 tgetgcaatgcttccctgcttagatgaagaacatacaaacacgcatggctctaaa 1258  
|||||  
Db 504 tgetgcaatgcttccctgcttagatgaagaacatacaaacacgcatggctctaaa 560

RESULT 9  
AAH99356  
ID AAH99356 standard; cDNA; 729 BP.  
XX  
AC AAH99356;  
XX  
DT 19-MAY-2000 (first entry)  
XX  
DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:2335.  
XX  
KW Human; colon cancer; tumour; diagnosis; gene expression product;  
KW probe; detection; cancerous state; metastasis; identification;  
KW breast cancer; oestrogen receptor-positive breast cancer; therapy;  
KW oestrogen receptor-negative breast cancer; lung cancer; ss.  
XX  
OS Homo sapiens.

XX WO9958675-A2.  
PN  
XX  
PD 18-NOV-1999.  
XX  
XX 13-MAY-1999; 99WO-US10602.  
XX  
XX 14-MAY-1998; 98US-0085426.  
PR 15-MAY-1998; 98US-0085537.  
PR 15-MAY-1998; 98US-0085690.  
PR 21-OCT-1998; 98US-0105234.  
PR 27-OCT-1998; 98US-0105877.  
XX  
XX (CHIR ) CHIRON CORP.  
PA (HYSE-) HYSEQ INC.  
XX  
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;  
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;  
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;  
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;  
XX  
XX WPI; 2000-126369/11.  
DR  
XX  
XX Polynucleotide library used to determine cancerous states of mammalian  
PT cells -  
PT  
XX  
PS Claim 1; Page 924-925; 1097pp; English.  
XX  
CC AAA00010 to AAA02716 represent polynucleotides isolated from cDNA  
CC libraries constructed from human colon cancer cell lines. The present  
CC invention also describes a method of detecting differentially expressed  
CC genes correlated with a cancerous state of a mammalian cell, comprising  
CC detecting at least one differentially expressed gene product in a test  
CC sample derived from a cell suspected of being cancerous, where detection  
CC of the differentially expressed gene product is correlated with a  
CC cancerous state of the cell from which the test sample was derived.  
CC The polynucleotides sequences can be used in a method for detecting  
CC differentially expressed genes correlated with a cancerous state of a  
CC mammalian cell. The polynucleotides can also be used as probes for  
CC detecting and mapping related genes. They can be used in diagnosis and  
CC prognosis of diseases and disorders (e.g. identification of  
CC pre-metastatic or metastatic cancerous states, stages of cancer, or  
CC responsiveness of cancer to therapy). This is particularly for breast  
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-  
CC negative breast cancer, lung cancer, and colon cancer.  
XX  
SQ Sequence 729 BP; 232 A; 178 C; 163 G; 143 T; 13 other;

Query Match 21.1%; Score 479; DB 21; Length 729;  
Best Local Similarity 99.8%; Pred. No. 3.9e-219;  
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1445 tgtgcacatcactgactggtaccacactctcattcactggtcgaagacagattgatga 1504  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
47 tgtgcacatcactgactggtaccacactctcattcactggtcgaagacagattgatga 106  
QY 1505 ggacattcaactagatgctatgatctatgtggagaccataagtgagggtcttcgcacc 1564  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
107 ggacattcaactagatgctatgatctggtggagaccataagtgagggtcttcgcacc 166  
QY 1565 ccgagtagatatatttgcaataacatgaccccatatcacaccaggccaaaatggctcctg 1624  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
167 ccgagtagatatatttgcaataacatgaccccatatcacaccaggccaaaatggctcctg 226  
QY 1625 ggcagcaggctatggatctggaacactgcaatccagtcagccatcagatgagcagcactg 1684  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
227 ggcagcaggctatggatctggaacactgcaatccagtcagccatcagatgagcagcactg 286  
QY 1685 gaaattgcttacaggaatactctggtacagcagactgggtccccctcagttttcagcaa 1744  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
287 gaaattgcttacaggaatactctggtacagcagactgggtccccctcagttttcagcaa 346

QY 1745 cctgggaccgaaccggtggcacaatgaacgcatcacctcgtcaactggcaaaagtgtatg 1804  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
347 cctgggaccgaaccggtggcacaatgaacgcatcaccttgtcaactggcaaaagtgtatg 406  
QY 1805 gcttttcaacatcacagccgaccccatatgagagggtggacctatctaacaggtatccagg 1864  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
407 gcttttcaacatcacagccgaccccatatgagagggtggacctatctaacaggtatccagg 466  
QY 1865 aatcgtgaagaagctccttacggaggtctcacagttcaacaaaactgacgtgccggtcag 1924  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
467 aatcgtgaagaagctccttacggaggtctcacagttcaacaaaactgacgtgccggtcag 526  
QY 1925 gtatccccccaaagaccccaagaagtaaccctaggtcctcaatggagggtct 1974  
Db |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
527 gtatccccccaaagaccccaagaagtaaccctaggtcctcaatggagggtct 576

RESULT 10  
AAS41205  
ID AAS41205 standard; cDNA; 629 BP.  
XX  
AC AAS41205;  
XX  
DT 17-DEC-2001 (first entry)  
XX  
DE cDNA encoding novel human enzyme polypeptide #421.  
XX  
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;  
KW ligase; hyperproliferative disorder; immunodeficiency disorder;  
KW autoimmune disorder; neurological disorder; metabolic disorder;  
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;  
KW blood-related disorder; infectious disorder; gene therapy; cytostatic;  
KW anti arthritic; nephrotropic; anticoagulant; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200155301-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01239.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 11-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 01-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-465566/50.  
DR P-PSDB; AAU23335.  
XX  
PT Novel polypeptides and polynucleotides useful for diagnosing,  
PT preventing, treating neural, immune system, muscular, reproductive,  
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous  
PT diseases -  
XX  
PS Claim 4; SEQ ID No 431; 1180pp; English.  
XX  
CC The present invention relates to the isolation of novel human enzyme  
CC polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences  
CC encoding them. The enzyme polypeptides of the invention may comprise the  
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,  
CC isomerases or ligases. The sequences of the invention are useful in the  
CC diagnosis, treatment, prevention and/or prognosis of a wide range of  
CC disorders including hyperproliferative disorders (e.g. cancer),  
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders  
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),  
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders  
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),  
CC blood-related disorders (e.g. haemophilia), reproductive disorders  
CC (e.g. infertility) and infectious disorders (e.g. Influenza). The  
CC polynucleotides of the invention can also be used in gene therapy.  
CC AAU40785-AAU41684 represent cDNA sequences encoding for the novel human  
CC enzyme polypeptides of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcc\_sequences.  
XX  
SQ Sequence 629 BP; 189 A; 151 C; 124 G; 162 T; 3 other;

Query Match 19.2%; Score 434; DB 22; Length 629;  
Best Local Similarity 99.8%; Pred. No. 1.4e-197;  
Matches 554; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 589 gagatgtgggttacacgagctgagattaaaacacactactcttgacaaagctcgctgcg 648  
|||||  
Db 14 gagatgtgggttacacgagctgagattaaaacacactactcttgacaaagctcgctgcg 73  
|||||  
QY 649 aaggagttaaactggagaaactactatgtctccagcctatttgacacaccatccaggagtcagt 708  
|||||  
Db 74 aaggagttaaactggagaaactactatgtctccagcctatttgacacaccatccaggagtcagt 133  
|||||  
QY 709 ttattactggaagtatcagatatacacacgcggaacttcaacattctctatcataagacctaacc 768  
|||||  
Db 134 ttattactggaagtatcagatatacacacgcggaacttcaacattctctatcataagacctaacc 193  
|||||  
QY 769 aaccacactgtttactctctgacaaatgcacccctacccacagaaactgaagaggttgat 828  
|||||  
Db 194 aaccacactgtttactctctgacaaatgcacccctacccacagaaactgaagaggttgat 253  
|||||  
QY 829 attcaacgcatatgctcgaaatggcacttgggtttttacagaaaagaatgcacgccca 888  
|||||  
Db 254 attcaacgcatatgctcgaaatggcacttgggtttttacagaaaagaatgcacgccca 313  
|||||  
QY 889 ccagaagaggattgtatacctttttt-ggtcccttttgggaagtggggattactatataca 947  
|||||  
Db 314 ccagaagaggattgtataccttttttgggtcccttttgggaagtggggattactatataca 373  
|||||  
QY 948 cactcaaatgtgacagtcctcggtggtggtggtgctatgacttgtatgaaaacgacaatgct 1007  
|||||  
Db 374 cactcaaatgtgacagtcctcggtggtggtggtgctatgacttgtatgaaaacgacaatgct 433  
|||||  
QY 1008 gcttgggactatgacaatggcatatactctccacacagatgtacactcagagagtagacgaa 1067  
|||||  
Db 434 gcttgggactatgacaatggcatatactctccacacagatgtacactcagagagtagacgaa 493  
|||||  
QY 1068 attcttagcttccataaaccacaaagcctatatattttatatattgctctatacaagctgtt 1127  
|||||  
Db 494 attcttagcttccataaaccacaaagcctatatattttatatattgctctatacaagctgtt 553  
|||||  
QY 1128 cattcaccactgcaa 1142  
|||||  
Db 554 cattcaccactgcaa 568  
|||||

RESULT 11  
AAA00278  
ID AAA00278 standard; cDNA; 294 BP.  
XX  
AC AAA00278;  
XX  
DT 19-MAY-2000 (first entry)  
XX  
DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:269.  
XX  
KW Human; colon cancer; tumour; diagnosis; gene expression product;  
KW probe; detection; cancerous state; metastasis; identification;  
KW breast cancer; oestrogen receptor-positive breast cancer; therapy;  
KW oestrogen receptor-negative breast cancer; lung cancer; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO9958675-A2.  
XX  
PD 18-NOV-1999.  
XX  
PF 13-MAY-1999; 99WO-US10602.  
XX  
PR 14-MAY-1998; 98US-0085426.  
PR 15-MAY-1998; 98US-0085537.  
PR 15-MAY-1998; 98US-0085696.  
PR 21-OCT-1998; 98US-0105234.  
PR 27-OCT-1998; 98US-0105877.  
XX  
PA (CHIR ) CHIRON CORP.  
FA (HYSE-) HYSEQ INC.

XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;  
PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;  
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;  
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;  
XX WPI; 2000-126369/11.  
XX Polynucleotide library used to determine cancerous states of mammalian  
PI cells -  
XX  
PS Claim 1; Page 244-245; 1097pp; English.  
XX  
CC AAA00010 to AAA02716 represent polynucleotides isolated from cDNA  
CC libraries constructed from human colon cancer cell lines. The present  
CC invention also describes a method of detecting differentially expressed  
CC genes correlated with a cancerous state of a mammalian cell, comprising  
CC detecting at least one differentially expressed gene product in a test  
CC sample derived from a cell suspected of being cancerous, where detection  
CC of the differentially expressed gene product is correlated with a  
CC cancerous state of the cell from which the test sample was derived.  
CC The polynucleotides sequences can be used in a method for detecting  
CC differentially expressed genes correlated with a cancerous state of a  
CC mammalian cell. The polynucleotides can also be used as probes for  
CC detecting and mapping related genes. They can be used in diagnosis and  
CC prognosis of diseases and disorders (e.g. identification of  
CC pre-metastatic or metastatic cancerous states, stages of cancer, or  
CC responsiveness of cancer to therapy). This is particularly for breast  
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-  
CC negative breast cancer, lung cancer, and colon cancer.  
XX  
SQ Sequence 294 BP; 79 A; 46 C; 51 G; 102 T; 16 other;  
  
Query Match 2.4%; Score 55; DB 21; Length 294;  
Best Local Similarity 100.0%; Pred. No. 5.4e-16;  
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1481 actgctgaagacagattgatgagacattcaactagatgctatgatctctgg 1535  
Db 65 actgctgaagacagattgatgagacattcaactagatgctatgatctctgg 119  
|||||  
  
RESULT 12  
AAA02370  
ID AAA02370 standard; cDNA; 1082 BP.  
XX  
AC AAA02370;  
XX  
DT 19-MAY-2000 (first entry)  
XX  
DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:2361.  
XX  
KW Human; colon cancer; tumour; diagnosis; gene expression product;  
KW probe; detection; cancerous state; metastasis; identification;  
KW breast cancer; oestrogen receptor-positive breast cancer; therapy;  
KW oestrogen receptor-negative breast cancer; lung cancer; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO9958675-A2.  
XX  
PD 18-NOV-1999.  
XX  
PF 13-MAY-1999; 99WO-US10602.  
XX  
PR 14-MAY-1998; 98US-0085426.  
PR 15-MAY-1998; 98US-0085537.  
PR 15-MAY-1998; 98US-0085696.  
PR 21-OCT-1998; 98US-0105234.  
PR 27-OCT-1998; 98US-0105877.  
XX  
PA (CHIR ) CHIRON CORP.



```
PR 22-MAY-1998; 98US-0086392.
PR 22-MAY-1998; 98US-0086414.
PR 22-MAY-1998; 98US-0086430.
PR 22-MAY-1998; 98US-0086486.
PR 28-MAY-1998; 98US-0087098.
PR 28-MAY-1998; 98US-0087106.
PR 28-MAY-1998; 98US-0087208.
PR 30-JUL-1998; 98US-0094651.
PR 11-SEP-1998; 98US-0100038.
XX (GETH ) GENENTECH INC.
PA Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
XX WPI; 1999-551358/46.
XX
XX New secreted and transmembrane polypeptides and their polynucleotides,
PT useful for treating blood coagulation disorders, cancers and cellular
PT adhesion disorders -
XX
XX Example 19; Page 197; 530pp; English.
PS
CC The present invention describes secreted and transmembrane polypeptides
CC and their polynucleotides. The nucleotide sequences are useful as
CC sources of probes, primers, for chromosome mapping, and for generation
CC of antisense sequences. They can also be used to create transgenic
CC animals. The proteins can be used to treat a variety of diseases and
CC disorders, depending on their function. Diseases that may be treated
CC include blood coagulation disorders, cancers and cellular adhesion
CC disorders. They may also be used to raise antibodies. AA233891 to
CC AA234338, and AA41685 to AA41774 represent polynucleotide and
CC polypeptide sequence given in the exemplification of the present
CC invention.
XX
SQ Sequence 53 BP; 14 A; 13 C; 14 G; 11 T; 1 other;

Query Match 1.4%; Score 32; DB 20; Length 53;
Best Local Similarity 100.0%; Pred. No. 5.6e-05;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 795 gccaccctacctcagaaactgaaggaggttg 826
Db 1 gccaccctacctcagaaactgaaggaggttg 32

RESULT 14
AAC78675
ID AAC78675 standard; DNA; 53 BP.
XX
AC AAC78675;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human PRO708 hybridisation probe SEQ ID NO:117.
XX
KW Human; secreted protein; transmembrane protein; PRO; EST; cytostatic;
KW expressed sequence tag; detection; cancer; PCR primer; probe; ss.
XX
OS Homo sapiens.
XX
FN WO200053756-A2.
XX
PD 14-SEP-2000.
XX
PF 18-FEB-2000; 2000WO-US04341.
XX
PR 08-MAR-1999; 99WO-US05028.
PR 12-MAR-1999; 99US-0123957.
PR 29-MAR-1999; 99US-0126773.
PR 21-APR-1999; 99US-0130232.
PR 28-APR-1999; 99US-0131445.
PR 14-MAY-1999; 99US-0134287.
```

```
PR 23-JUN-1999; 99US-0141037.
PR 26-JUL-1999; 99US-0145698.
PR 29-OCT-1999; 99US-0162506.
PR 30-NOV-1999; 99WO-US28313.
PR 02-DEC-1999; 99WO-US28551.
PR 02-DEC-1999; 99WO-US28565.
PR 16-DEC-1999; 99WO-US30095.
PR 30-DEC-1999; 99WO-US31243.
PR 30-DEC-1999; 99WO-US31274.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00277.
PR 06-JAN-2000; 2000WO-US00376.
XX (GETH ) GENENTECH INC.
XX
XX Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
PI Ferrara N, Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;
PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
PI Kljavin IJ, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;
PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
XX WPI; 2000-611443/58.
XX
XX Novel PRO polypeptides and polynucleotides used in detection methods,
PT to target bioactive molecules to specific cells, and to modulate
PT cellular activities -
XX
XX Example 19; Page 250; 536pp; English.
XX
CC AAC78458 to AAC78599 represent polynucleotide and EST (expressed
CC sequence tag) sequences which encode secreted or transmembrane PRO
CC polypeptides. The PRO polynucleotides and polypeptides have cytostatic
CC activity. The polynucleotides and polypeptides can be used for detecting
CC the presence of PRO polypeptides in samples, for linking bioactive
CC molecules to cells and for modulating biological activities of cells,
CC using the polypeptides for specific targeting. The polypeptide targeting
CC can be used to kill the target cells, e.g. for the treatment of cancers.
CC The polypeptide pairs provide specific targeting of bioactive molecules
CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
CC the isolation of the PRO polynucleotide sequences.
XX
SQ Sequence 53 BP; 14 A; 13 C; 14 G; 11 T; 1 other;

Query Match 1.4%; Score 32; DB 21; Length 53;
Best Local Similarity 100.0%; Pred. No. 5.6e-05;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 795 gccaccctacctcagaaactgaaggaggttg 826
Db 1 gccaccctacctcagaaactgaaggaggttg 32

RESULT 15
AAK63748
ID AAK63748 standard; cDNA; 774 BP.
XX
AC AAK63748;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen encoding CDNA SEQ ID NO:8808.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ss.
XX
OS Homo sapiens.
XX
FN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
```



XX PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0218290.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 12-SEP-2000; 2000US-0232081.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249219.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-483426/52.  
DR P-PSDB; AAM90967.  
XX  
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,

PT useful for preventing, diagnosing and/or treating cancers and  
PT metastasis -

XX

PS Claim 1; SEQ ID NO 8808; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I) amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (I) by expressing inactive proteins or to supplement the patients' own production of (I). Additionally, (I) polynucleotides may be used to produce the secreted (I) by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) proteins and polynucleotides may be used to prevent, diagnose and treat immune/haematopoietic-related diseases, especially cancers and cancer metastases of haematopoietic-derived cells. AAK64703 to AAK87694 represent human immune/haematopoietic antigen genomic sequences from the present invention. AAK54942 to AAK54950 and AAK82169 represent sequences used in the exemplification of the present invention.

SQ Sequence 774 BP; 215 A; 194 C; 189 G; 173 T; 3 other;

### Query Match

Query Match

Best Local Similarity 100.0%; Pred. No. 0.043;

Qy 1 cacgcgtccgccccacgcgtccgtgga 26

7

Db 4 cacggtccgcccacgcgtccgtgga 29

Search completed: July 23, 2002, 19:01:42  
Job time: 11368 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 15:12:04 ; Search time 4094.62 Seconds  
(without alignments)  
11580.936 Million cell updates/sec

Title: US-09-495-823-8  
Perfect score: 2266  
Sequence: 1 caccgctcgccacgcgtc.....tgcacctggtgccgaattc 2266

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htgo\_inv.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
------------	-------	-------	-------	--------	----	-------------

RESULT 1

AX206967

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

CDS

AX206967 2253 bp . DNA linear PAT 30-AUG-2001  
Sequence 8 from Patent WO0155411.  
AX206967  
AX206967.1 GI:15394724  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 2253)  
Glucksmann, M.A., Williamson, M., Rudolph-Owen, L.A. and Tsai, F.Y.  
Human sulfatases  
Patent: WO 0155411-A 8 02-AUG-2001;  
Millennium Pharmaceuticals, Inc. (US)  
Location/Qualifiers  
1..2253  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
324..2123  
/note="unnamed protein product"  
/codon\_start=1

ALIGNMENTS

1	1877	82.8	2253	6	AX206967	Sequence
2	1549	68.4	1800	6	AX206973	Sequence
c 3	1207	53.3	152861	9	AC104779	Homo sapi
c 4	1207	53.3	165147	2	AC021342	Homo sapi
c 5	630	27.8	157043	2	AC013692	Homo sapi
c 6	630	27.8	165147	2	AC021342	Homo sapi
c 7	630	27.8	175645	2	AC105414	Homo sapi
c 8	263	11.6	1871	9	AK027201	Homo sapi
c 9	99	4.4	176341	2	AC105383	Homo sapi
c 10	53	2.3	232951	2	AC091322	Mus muscu
c 11	29	1.3	122992	2	AC098604	Rattus no
c 12	26	1.1	927	9	BC009528	Homo sapi
c 13	26	1.1	2003	10	BC012659	Mus muscu
c 14	25	1.1	28735	2	AC108339	Rattus no
c 15	25	1.1	38866	3	CBRG36E21	Caenorhab
c 16	25	1.1	191744	2	AC044865	Homo sapi
c 17	25	1.1	211652	9	AP001838	Homo sapi
c 18	24	1.1	1603	10	RATARSB	D49434 Rat ARSB mr
c 19	24	1.1	2491	10	BC021406	Mus muscu
c 20	24	1.1	3695	10	BC021427	Mus muscu
c 21	24	1.1	118722	2	AP000584	Homo sapi
c 22	24	1.1	136503	2	AC068939	Homo sapi
c 23	24	1.1	145718	2	AC091369	Rattus no
c 24	24	1.1	153400	2	AC106436	Rattus no
c 25	24	1.1	153704	9	AC008567	Homo sapi
c 26	24	1.1	153882	2	AC011256	Homo sapi
c 27	24	1.1	161591	2	AP001142	Homo sapi
c 28	24	1.1	170537	2	AC073548	Homo sapi
c 29	24	1.1	173435	2	AP002386	Homo sapi
c 30	24	1.1	174484	9	AP002815	Homo sapi
c 31	24	1.1	176852	2	AC097993	Rattus no
c 32	24	1.1	180672	9	AC011451	Homo sapi
c 33	24	1.1	191919	9	AC016770	Homo sapi
c 34	24	1.1	219366	9	AP001922	Homo sapi
c 35	23	1.0	482	6	AR128083	Sequence
c 36	23	1.0	1216	6	AX277154	Sequence
c 37	23	1.0	1540	9	HS420512	Homo sapi
c 38	23	1.0	2954	10	BC021508	Mus muscu
c 39	23	1.0	3451	5	AF072455	Xenopus l
c 40	23	1.0	4096	9	BC015055	Homo sapi
c 41	23	1.0	4858	6	AX347398	Homo sapi
c 42	23	1.0	4858	6	AX349119	Sequence
c 43	23	1.0	4918	10	BC021490	Mus muscu
c 44	23	1.0	17514	2	AC012851	Drosophil
c 45	23	1.0	20280	9	AL671136	Human DNA

/protein\_id="CAC60191.1"  
/db\_xref="GI:15394725"  
/translation="MAPRGAGHPPPPPOACVCPGKMLAMGALAGRWILCLLTGYLV  
SWGQALEEEEGALLQAQGEKLEPSTSTSOPLHILFILDQDQFQVDRVGYHSGSEIKPT  
LDKLAEGVKLENYVQPICTPERSQFITKQIQLHTGLQHSIIIRPTOPNCLPDNATL  
PQKLKEVGYTHVMGWKHLFGYRKCPMTRRGEDTFEGLSGSDYTHYKCDSPGMC  
GYDLYENDNAWDYDNGIYSTQMYTORVOQILASHNPKPIFLYIAQVHSPLOAPG  
RYFEHRYININIRRYAAILMSCLIDEAINNVIALKTYGYNNIIISYSDNGQOPTA  
GDSNPLRGSKGYWEGGIRAVGFVHSPLLKNKGTVCVELVHTDTPPTLISLAEGOI  
DEDIQLDGYDIMEFISBGLSPRVDILHNIDPIYTKAKNGSWAAGYGIWNTAQSAIR  
VQHWKLLTGPNFGSDVMYPPQSFNLGNRNWHERITLSTGKSWLFNITADPIFERVDL  
SNRYPGIVKLLRLRLOSFNKTPVPRYPKDPSPRLNGVWMPGPKYKEETKKKKPSK  
NQAEKKKKKKKKKKOQKAVSGSTCHSGVTG"  
BASE COUNT 624 a 535 c 554 g 540 t  
ORIGIN

Query Match 82.8%; Score 1877; DB 6; Length 2253;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 2247; Conservative 0; Mismatches 5; Indels 1; Gaps 1;  
  
Qy 1 caecggtccgcacgagtcgctggagagataataactttttttttttttttttttttt 60  
Db 1 CACGGCTCCGCCACGGCTCCGCGGAGATATAACTTTTCTTTTCTTTTCTTGGT 60  
  
Qy 61 ggaagctgcttagggagggggagggagagaaagtgaatgctgctggagaagcgc 120  
Db 61 GGAAGCTGCTTAGGAGGGGGAGGAGGAGGAGAAAGTGAANTGCTGTGAGAGAGCG 120  
  
Qy 121 agccctctgttttccggagtcocattccatttaagccatcacttctggaagataaagt 180  
Db 121 AGCCCTCTTCTTCTCGGGAGTCCCATCCATTAAAGCCATCACTTCTGGAAGATTAAGT 180  
  
Qy 181 tctggacatggtcaacgtagaggagagagagaaattcttccaggtggagagcttca 240  
Db 181 TGTGGACATGCTGACAGTGAAGAGGAGAGAGGAGATTTCTTGCCAGGTGGAGAGTCTTCA 240  
  
Qy 241 ccgtctgttgggtgcattgtgcgcgcgcacscggcgggcgqggtgttctccgcgtgg 300  
Db 241 CCGTCTGTTGGTGCATGTGTGCCCCGACGCGGCGGGCGGTGTTCTCCGCGTGG 300  
  
Qy 301 agtctaacctgggacctgagtgatggtgtcccgaggggtgtgcgggggcaatccgcgc 360  
Db 301 AGTCTCACTGGGACCTGAGTGAATGGCTCCACGGGGGTGTGCGGGGCATCCGCTCCGC 360  
  
Qy 361 cttctccacagcgtgtctgtcctggaagatgctagcaaatggggcgctgcagagat 420  
Db 361 CTTCTCCACAGCCCTGTGTCTCTGGAAGATGCTAGCAATGGGGCGCTGGCAGGAT 420  
  
Qy 421 tctggatcctctgctcctcacttatggttacctgtccttggggcgagccttagaagagg 480  
Db 421 TCTGGATCCTCTGCTCTCACTATGTTACCTGTCTCTGGGGCAGGCGCTTAGAGAGG 480  
  
Qy 481 aggaagaaggggccttactagctcaagctgagagaaactagagcccgagcaacttcca 540  
Db 481 AGGAAGAAGGGGCGCTTACTAGCTCAAGCTGAGAGAAACTAGAGCCAGCAACTTCCA 540  
  
Qy 541 cctccagcccatctcatttccatctagcggatgatcaggagatttagagatgggtt 600  
Db 541 CCTCCAGCCCATCTCATTTTCTAGCGGATGATCAGGGATTTTAGAGATGTGGTT 600  
  
Qy 601 accacggatctgagatataaacacactactcttgacaagctcgctgcggaaggagttaaac 660  
Db 601 ACCAGGATCTGAGATTAACACCTACTCTTGACAGCTCGCTGCCGAGAGGATTAAC 660  
  
Qy 661 tgggaactactatgtccagcctatttgcacaccatccaggagtcagtttattactggaa 720  
Db 661 TGGAGAACTACTATGTCTCCAGCCTATTTGCACACCATCCAGGAGTCAGTTTATTACTGAA 720  
  
Qy 721 agtatcagatcacacggagctcaacattctatcaatgaagcctaccaccccaacttt 780  
Db 721 AGTATCAGATACACCGGACTTCAACATTCTATCATAAGACCTTACCACCACTACTTT 780

Qy 781 tacctctggacaagtccaccctacactcagaaactgaaagaggttgatattcaacgcata 840  
Db 781 TACCTCTGGAACAATGCCACCTACCTCAGAACTGAAGAGGTTTGGATATTCAACGCATA 840  
  
Qy 841 tggctggaaaaatggcacttgggttttttttacaaaaaagaatgcatgccaccagaaagat 900  
Db 841 TGGTTCGAAAAATGGCAGTGGGTTTTTACAAAAAAGATGCGTCCACCAGAAAGGAT 900  
  
Qy 901 ttgataccttttttgggttcccttttgggaagtgggagttactatcacacactacaagt 960  
Db 901 TTGATACCTTTTTTGGTTCCTTTTGGGAAGTGGGGATTACTATACACACTACAAGTATG 960  
  
Qy 961 acagtccctgggagtgctggctatgacttgatgaaaacgacaatgctgcctgggactatg 1020  
Db 961 ACAGTCTTGGGATGTGCTGCTATGACTTGTATGAAAACGACAATGCTGCTGGGACTATG 1020  
  
Qy 1021 acaatggcatactactccacacagatgtacactcagagagtcacagcaaatcttagtctccc 1080  
Db 1021 ACAATGGCATATACTCCACACAGATGTACACTCAGAGAGTACAGCAAAATCTTAGCTTCC 1080  
  
Qy 1081 ataacccacaaaagcctatttttatattgctctatcaagctgttctcattcaccactgc 1140  
Db 1081 ATAAACCCACAAAAGCCTATATTTTATATATGCTTATCAAGCTGTTTCTTCACTCACCAC 1140  
  
Qy 1141 aagctccctggcaggtatttcgaacactaccgatccattatcaatacaacagagagagat 1200  
Db 1141 AAGCTCTTGGCAGGTATTTTCGAACACTACCGATCCATTATCAACATAAACAGAGAGAT 1200  
  
Qy 1201 atgctgcatacgttctcctgcttagatgaagcaatcaacaacgtgcattggctctaaaga 1260  
Db 1201 ATGTGCGCATGCTTTTCTCTTAGATGAAGCAATCAACAACGTGACATTTGGTCTCTAAAGA 1260  
  
Qy 1261 ctatggtttctataacaacagcattatcatttactcttcagataatggtggccagccta 1320  
Db 1261 CTTATGTTTCTTATAAACACAGCATATATCTTTACTCTTCAGATTAATGGTGGCCAGCCTA 1320  
  
Qy 1321 cggcaggaggagtaactggcctctcagaggtagcaaggaacaatatgggaagagagga 1380  
Db 1321 CGGCAGGAGGAGTAACTGGCTCTCAGAGGTAGCAAAAGGAACAATATGGGAAGAGGGA 1380  
  
Qy 1381 tccgggctgtagggtttgtgcatagcccaactctgaaaaaagaaggaacagtggttaagg 1440  
Db 1381 TCCGGGCTGTAGGGTTTGTGCATAGCCCCACTTCTGAAAAACAAGGAACAAGTGTGTAAGG 1440  
  
Qy 1441 aactgtgcacatcaactggttaccacactctatttctcactgctgctgaagacagatcg 1500  
Db 1441 AACTGTGCACATCACTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500  
  
Qy 1501 atgaggaattcaacatagatggtatgatatctgggagaccataagtgaggtcttctcgt 1560  
Db 1501 ATGAGGACATTCACATAGATGGCTATGATATCTGGGAGACCATAAGTGAGGCTCTTCGCT 1560  
  
Qy 1561 caccgcgagtagataatttgcataaactgaccccatatacaccagaagcgaataaagt 1620  
Db 1561 CACCCGAGTAGATATTTTGCATAACATTTAGCCCCATATACACCAGGCAAAATTTGGCT 1620  
  
Qy 1621 cctgggcagcaggtatggatctggaacactgcgaatccagtcagcccatcagagtcagc 1680  
Db 1621 CCTGGGCAGCAGGCTATGGGATCTGGAACACTGCAATCCAGTCAGCCATCAGAGTGCAGC 1680  
  
Qy 1681 actggaattgtctacagaaatcctggctacagcactgggtccccctcagctcttca 1740  
Db 1681 ACTGGAATTTGCTTACAGGAAATCTGGCTACAGCAGTGGGTCCCCCTCAGTCTTTCA 1740  
  
Qy 1741 gcaacctgggaccgaacccgggtggcaaatgaacgaatcaacctcgtaactggcaaaagt 1800  
Db 1741 GCAACTGGGACCGAACCCGGTGGCACAATGAACCGGATCACCTTGTCACTGCGCAAAAGTG 1800  
  
Qy 1801 tatggcttttcaacatcacagcgaccctatgatgaggttgagccttacttaacaggtacc 1860  
Db 1801 TATGGCTTTTCAACATCACAGCCGACCCCATATGAGAGGGTGGACCTATCTTAACAGGTATC 1860  
  
Qy 1861 caggaaatcgtagaagaagctcctacgaggtctctcagttcaacaaaaactgcagtgccg 1920



Db	1321	TGGAACTGCAATCCAGTACGCCATCAGAGTGACGACTGGAATTCCTTACAGGAAAT	1380
Qy	1704	cctggctacagcgactgggtccccctcagctctttcagcaacctgggaccgaaccggtgg	1763
Db	1381	CCTGGCTACACGGACTGGGTCCCCCTCAGTCTTTCAGCACACCTGGGACCGGTGG	1440
Qy	1764	cacaatgaacggatcacctcgctcaactggcaaaagtgtatggcttttcaacatcacagcc	1823
Db	1441	CACAATGAACGGATCACCTTCTCAACTGGCAAAAGTGATGGCTTTTCAACATCACAGCC	1500
Qy	1824	gaccatatgagaggggtggaacctatctaaacaggtatccaggaatctgtaagaagctcta	1883
Db	1501	GACCCATATGAGAGGGTGGACCTTCTAACAGGTATCCAGGAATCGTGAAGAAGCTCCTA	1560
Qy	1884	cgaggctctcacagttcaacaaactgcagtccggtcaggtatccccccaaagacccc	1943
Db	1561	CGGAGGCTCTCACAGTTCAACAAACTCGAGTGGCGGTGAGGTATCCCCCCAAAGACCCC	1620
Qy	1944	agaagtacacctaggtcctaattggagggtct	1974
Db	1621	AGAAGTAAACCTAGGCTCAATGGAGGGGTCT	1651
RESULT	3		
AC104779/c			
LOCUS	AC104779	152861 bp	DNA linear PRI 09-FEB-2002
DEFINITION	Homo sapiens chromosome 4 clone RP11-26P13, complete sequence.		
ACCESSION	AC104779		
VERSION	AC104779.4	GI:18642899	
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	Waterston, R.H.		
JOURNAL	The sequence of Homo sapiens clone		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 152861)		
TITLE	Waterston, R.H.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (21-FEB-2001) Genome Sequencing Center, Washington		
AUTHORS	University School of Medicine, 4444 Forest Park Parkway, St. Louis,		
TITLE	MO 63108, USA		
JOURNAL	3 (bases 1 to 152861)		
REFERENCE	Waterston, R.H.		
AUTHORS	Direct Submission		
TITLE	Submitted (09-FEB-2002) Genome Sequencing Center, Washington		
JOURNAL	University School of Medicine, 4444 Forest Park Parkway, St. Louis,		
COMMENT	MO 63108, USA		
	On Feb 9, 2002 this sequence version replaced gi:18482303.		
FEATURES			
source	Center: Washington University Genome Sequencing Center		
	Center code: WUGSC		
	Web site: <a href="http://genome.wustl.edu/gsc/index.shtml">http://genome.wustl.edu/gsc/index.shtml</a>		
	Contact: <a href="mailto:submissions@wustl.edu">submissions@wustl.edu</a>		
	----- Project Information -----		
	Center project name: H_NH0026P13		
	----- Location/Qualifiers -----		
	1. .152861		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="4"		
	/clone="RP11-26P13"		
BASE COUNT	48870 a 27681 c 28505 g 47805 t		
ORIGIN			
Query Match	53.3%	Score 1207;	DB 9; Length 152861;
Best Local Similarity	99.7%	Pred. No. 0;	
Matches 1527; Conservative	0; Mismatches	4; Indels	1; Gaps



```
misc_feature      /note="assembly_fragment"  
10980..15646  
/note="assembly_fragment"  
15747..25208  
/note="assembly_fragment"  
25309..33458  
/note="assembly_fragment"  
33559..46617  
/note="assembly_fragment"  
46718..60565  
/note="assembly_fragment"  
60666..92690  
/note="assembly_fragment"  
92791..122716  
/note="assembly_fragment"  
122817..149492  
/note="assembly_fragment"  
149593..165147  
/note="assembly_fragment  
clone_end:T7  
vector_side:right"  
BASE COUNT      52130 a 29805 c 30056 g 51853 t 1303 others  
ORIGIN  
  
Query Match      53.3%; Score 1207; DB 2; Length 165147;  
Best Local Similarity 99.7%; Pred. No. 0;  
Matches 1527; Conservative 0; Mismatches 4; Indels 1; Gaps 1;  
  
Qy 722 gatacagatacacaccggactcaacattctatcatcaagacctaccacccaactgttt 781  
Db 101880 GTATCAGATACACACCGGACATTCACATTCATCATAGACCTACCCCAACCAACTGT 101821  
  
Qy 782 acctctggacaatgccaccctacctcagaaactgaagaggttgatatcaacgcatat 841  
Db 101820 ACCTCTGGACAATGCCACCCTACCTCGAAGACTGAAGGAGGTGGATATTCAACGCAAT 101761  
  
Qy 842 ggtcggaaaaatggcacttgggtttttacagaaaagaaatgcatagccaccagaaagagatt 901  
Db 101760 GGTGCGAAAATGGCACTTGGGTTTTACAGAAAAGAAATGCATGCCACCACGAAGAGATT 101701  
  
Qy 902 tgataccttttttgggttcccttttgggaagtggggattactatatacacactacaatgtga 961  
Db 101700 TGATACCTTTTTTGGTTCCTTTTGGGAAGTGGGGATTACTATACACTACAAATGTGA 101641  
  
Qy 962 cagtcctgggagtgctggctatgactgtatgaaacagacaactgcctgggactatga 1021  
Db 101640 CAGTCCTGGGATGTGGCTATGACTTGTATGAAGACGACAATGCTGCTGGGACTATGA 101581  
  
Qy 1022 caatggcatatactccacacagatgtacactcagagagtacagcaaatcttagcttccca 1081  
Db 101580 CAATGGCATATATCCACACAGATGTACACTCAGAGAGTACAGCAAAATCTTAGCTTCCCA 101521  
  
Qy 1082 taaccccaacaagcctatatttttatattatgtcctatacaagctgttcaatccactgca 1141  
Db 101520 TAAACCCACAAGGCTATATTTTATATATATGCGCTATCAAGCTGTTCATTCAACACTGCA 101461  
  
Qy 1142 agctcctggcagggtatttcgaacactaccgataccattatcaacataaagagagagata 1201  
Db 101460 AGCTCCTGGCAGGATTTTCGAACACTACCGGATCCATTATCAACATTAACAGGAGGAGATA 101401  
  
Qy 1202 tgcctgccatgctttcctgcttagatgaagcaatcaacaactgcacattggctctaaagac 1261  
Db 101400 TGCTGCCATGCTTTCCTGCTTAGATGAAGCAATCAACAACGTGACATGGCTCTAAAGAC 101341  
  
Qy 1262 ttatggtttctataacaacagcatataatcttactcttcagataatgggtgccagcctac 1321  
Db 101340 TTATGGTTTCTATAAACACACATTATCATTTACTTCTTCAGATATGTTGGTCCAGCCTAC 101281  
  
Qy 1322 ggcaggaggagtaactggcctctcaaggtagcaaggaacatataggaaagagagatt 1381  
Db 101280 GGCAGGAGGAGTAACTGGCCCTCTCAGAGGTAGCAAGGAACATATTGGGAAGGAGGAT 101221
```

```
Qy 1382 ccgggctgtaggctttgtgcatagcccaacttctgaaaaacaaggaaacagtggtgtaagga 1441  
Db 101220 CCGGCTGTAGGCTTTGTGCATAGCCCACTTCTGAAAAACAGGAAACAGTGTGTAAAGGA 101161  
  
Qy 1442 acctgtgcacatcaactgactggttaaccacactctctattcactggctgaagacagattga 1501  
Db 101160 ACTTGTGCACATCACTGACTGGTACCCCACTCTCATTTTCACTGGCTGAAGACAGATTGA 101101  
  
Qy 1502 tgaggacattcaactagatggctatgatatctggagacccaatgaaggtcttcctgcctc 1561  
Db 101100 TGAGGACATTCAACTAGATGGCTATGATATCTGGGAGACCATAAGTAGGGGTCTTCGGCTC 101041  
  
Qy 1562 accccgagtagatatatttgcataaacttgaccatacacaccaaggcaaaaaatggctc 1621  
Db 101040 ACCCCGAGTAGATATTTTGCATAACATTTGACCCCATATACACCAAGGCAAAAATGGCTC 100981  
  
Qy 1622 ctgggcagcaggctatgggatactggaaacactgcaatccagtcagcca tcagagtgcagca 1681  
Db 100980 CTGGGCAAGCAGGCTATGGGATCTGGAACACTGCAATTCAGTCAGGCATCAGAGTCAGCA 100921  
  
Qy 1682 ctggaattgcttacagaaaatcctgctacagcagactgggtcccccctcagttcttcag 1741  
Db 100920 CTGGAATTGCTTACAGGAAATCCTGGCTACAGCACTGGGTGCCCTCTAGTCTTTTCAG 100861  
  
Qy 1742 caacctgggaccggaacccggtggcacaatgaacggatcacctcgtcaactggcaaaaagtgt 1801  
Db 100860 CAACCTGGGACCGGAACCGGTGGCACAAATGAACGGATCACTTGTCACTGCAAAAAGTGT 100801  
  
Qy 1802 atgcttttcaaatcacagccgacccaatatgagaggggtggacctatctacaaggtatcc 1861  
Db 100800 ATGCTTTTCAACATCACAGCCGACCCCATATGAGAGGGTGGACCTATCTAACAGGTATCC 100741  
  
Qy 1862 aggaatcgtggaagctcctacgaggtctcacagttctcaagttcaacaaaactgcagtcagg 1921  
Db 100740 AGGAATCCTGAAGAGCTCTTACGGAGCTCTCACAGTTTCAACAAAACCTGCAGTGCCTGGT 100681  
  
Qy 1922 caggtatcccccccaagaccaccagaagtaaccttaggtcattgaggggtctcaggggacc 1981  
Db 100680 CAGGTATCCCCCAAGACCCCAAGAGTAACCTTAGGCTCAATGGAGGGTCTGGGGACC 100621  
  
Qy 1982 atggtatagagagaacccaagaaagaaagccaaagcaaaaaatcaggctgagaaaaagca 2041  
Db 100620 ATGGTATAAAGAGAAACCAAGAAAAGAGCCAAAGCAAAAATCAGGCTGAGAAAAAGCA 100561  
  
Qy 2042 aaaaaaagc-aaaaaagaagaagaaacagacagaagcagctctcaggttcaacttcgca 2100  
Db 100560 AAAGAAAAGCAAAAAGAAAAGAAAGAAACACAGCAAGAACAGCTCTCAGGTTTCAACTTGCCA 100501  
  
Qy 2101 ttcagggttacttctggtgataagcacaaaatatttccctgtttgtggttaaaactttaatcagtt 2160  
Db 100500 TTCAGGTGTTACTTGTGGATAAGCACAAATATTTCCCTGTTTGGTTAAACTTTTAATCACTT 100441  
  
Qy 2161 cttattcttcatctgtttcctcaggttaaacagcaaaatttggctcgataataatcgtcgacc 2220  
Db 100440 CTTATCTTTCATCTGTTTCTTAGTAAACACAGCAAAATTTGGCTCGATAATATATCGTGGCC 100381  
  
Qy 2221 taagcgtcaggctgttttctcagctgtgtccac 2252  
Db 100380 TAAGCGTCAGGCTGTTTTCATGCTGTGCCAC 100349  
  
RESULT 5  
AC013692/c  
LOCUS  
DEFINITION Homo sapiens clone RP11-21L8, *** SEQUENCING IN PROGRESS ***, 45  
unordered pieces.  
ACCESSION AC013692  
VERSION AC013692.5 GI:11597077  
KEYWORDS HTG; HTGS_PHASE1.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```



REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

1 (bases 1 to 157043)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome, clone RP11-21L8  
Unpublished  
2 (bases 1 to 157043)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,  
Brown,A., Castile,A., Colangelo,M., Collins,S., Collymore,A.,  
Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,  
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,  
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Lehotzky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,  
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,  
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,  
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.  
Direct Submission

TITLE  
JOURNAL  
COMMENT

Submitted (13-NOV-1999), Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Dec 7, 2000 this sequence version replaced gi:10280848.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WtBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L4053  
Center clone name: 21\_L\_8  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 45 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 924: contig of 924 bp in length  
\* 925 1024: gap of 100 bp  
\* 1025 1800: contig of 776 bp in length  
\* 1801 1900: gap of 100 bp  
\* 1901 2800: contig of 900 bp in length  
\* 2801 2900: gap of 100 bp  
\* 2901 3628: contig of 728 bp in length  
\* 3629 3728: gap of 100 bp  
\* 3729 4582: contig of 854 bp in length  
\* 4583 4682: gap of 100 bp  
\* 4683 5763: contig of 1081 bp in length  
\* 5764 5863: gap of 100 bp  
\* 5864 6992: contig of 1129 bp in length  
\* 6993 7092: gap of 100 bp  
\* 7093 8691: contig of 1599 bp in length  
\* 8692 8791: gap of 100 bp  
\* 8792 10237: contig of 1446 bp in length  
\* 10238 10337: gap of 100 bp  
\* 10338 12115: contig of 1778 bp in length  
\* 12116 12215: gap of 100 bp  
\* 12216 14480: contig of 2265 bp in length  
\* 14481 14580: gap of 100 bp  
\* 14581 16200: contig of 1620 bp in length  
\* 16201 16300: gap of 100 bp  
\* 16301 17801: contig of 1501 bp in length  
\* 17802 17901: gap of 100 bp  
\* 17902 19560: contig of 1659 bp in length  
\* 19561 19660: gap of 100 bp  
\* 19661 21061: contig of 1401 bp in length  
\* 21062 21161: gap of 100 bp

\* 21162 22798: contig of 1637 bp in length  
\* 22799 22898: gap of 100 bp  
\* 22899 25224: contig of 2326 bp in length  
\* 25225 25324: gap of 100 bp  
\* 25325 27369: contig of 2045 bp in length  
\* 27370 27469: gap of 100 bp  
\* 27470 30262: contig of 2793 bp in length  
\* 30263 30362: gap of 100 bp  
\* 30363 32450: contig of 2088 bp in length  
\* 32451 32550: gap of 100 bp  
\* 32551 34415: contig of 1865 bp in length  
\* 34416 34515: gap of 100 bp  
\* 34516 37321: contig of 2806 bp in length  
\* 37322 37421: gap of 100 bp  
\* 37422 40476: contig of 3055 bp in length  
\* 40477 40576: gap of 100 bp  
\* 40577 50235: contig of 9659 bp in length  
\* 50236 50335: gap of 100 bp  
\* 50336 53457: contig of 3122 bp in length  
\* 53458 53557: gap of 100 bp  
\* 53558 56770: contig of 3213 bp in length  
\* 56771 56870: gap of 100 bp  
\* 56871 59626: contig of 2756 bp in length  
\* 59627 59726: gap of 100 bp  
\* 59727 62247: contig of 2521 bp in length  
\* 62248 62347: gap of 100 bp  
\* 62348 65709: contig of 3362 bp in length  
\* 65710 65809: gap of 100 bp  
\* 65810 69014: contig of 3205 bp in length  
\* 69015 69114: gap of 100 bp  
\* 69115 73176: contig of 4062 bp in length  
\* 73177 73276: gap of 100 bp  
\* 73277 78978: contig of 5702 bp in length  
\* 78979 79078: gap of 100 bp  
\* 79079 83397: contig of 4319 bp in length  
\* 83398 83497: gap of 100 bp  
\* 83498 87887: contig of 4390 bp in length  
\* 87888 87987: gap of 100 bp  
\* 87988 91530: contig of 3543 bp in length  
\* 91531 91630: gap of 100 bp  
\* 91631 97097: contig of 5467 bp in length  
\* 97098 97197: gap of 100 bp  
\* 97198 103021: contig of 5824 bp in length  
\* 103022 103121: gap of 100 bp  
\* 103122 108656: contig of 5535 bp in length  
\* 108657 108756: gap of 100 bp  
\* 108757 113320: contig of 4564 bp in length  
\* 113321 113420: gap of 100 bp  
\* 113421 120645: contig of 7225 bp in length  
\* 120646 120745: gap of 100 bp  
\* 120746 130194: contig of 9449 bp in length  
\* 130195 130294: gap of 100 bp  
\* 130295 137856: contig of 7562 bp in length  
\* 137857 137956: gap of 100 bp  
\* 137957 147125: contig of 9169 bp in length  
\* 147126 147225: gap of 100 bp  
\* 147226 154966: contig of 7741 bp in length  
\* 154967 155066: gap of 100 bp  
\* 155067 157043: contig of 1977 bp in length.

## FEATURES

source  
1. 157043  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-21L8"  
/clone\_lib="RPC1-11 Human Male BAC"  
1. 924  
/note="assembly\_fragment  
vector\_end:SP6  
vector\_side:left"  
1025..1800  
/note="assembly\_fragment"  
1901..2800  
/note="assembly\_fragment"

misc\_feature  
misc\_feature  
misc\_feature  
misc\_feature



```

Center clone name: 24_I-3
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 156719 bases at least Q40
Consensus quality: 160377 bases at least Q30
Consensus quality: 162195 bases at least Q20
Insert size: 170000; agarose-ftp
Insert size: 163847; sum-of-contigs
Quality coverage: 4.4 in Q20 bases; agarose-ftp
Quality coverage: 4.6 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
*
1      867: contig of 867 bp in length
      868 967: gap of 100 bp
      968 2133: contig of 1166 bp in length
      2134 2233: gap of 100 bp
      2234 3987: contig of 1754 bp in length
      3988 4087: gap of 100 bp
      4088 6971: contig of 2884 bp in length
      6972 7071: gap of 100 bp
      7072 10879: contig of 3808 bp in length
      10880 10979: gap of 100 bp
      10980 15646: contig of 4667 bp in length
      15647 15746: gap of 100 bp
      15747 25208: contig of 9462 bp in length
      25209 25308: gap of 100 bp
      25309 33458: contig of 8150 bp in length
      33459 33558: gap of 100 bp
      33559 46617: contig of 13059 bp in length
      46618 46717: gap of 100 bp
      46718 60565: contig of 13848 bp in length
      60566 60665: gap of 100 bp
      60666 92690: contig of 32025 bp in length
      92691 92790: gap of 100 bp
      92791 122716: contig of 29926 bp in length
      122717 122816: gap of 100 bp
      122817 149492: contig of 26676 bp in length
      149493 149592: gap of 100 bp
      149593 165147: contig of 15355 bp in length.

```

## FEATURES

[illegible]

## RESULTS

AC10541A/C

LOCUS

HTG 09-FEB-2002

DNA linear

175645 b1

AC1054

LOCUS

---

.....

0.46617

ure  
335

misc\_feat

```

DEFINITION Homo sapiens chromosome 4 clone RP11-219L11, WORKING DRAFT
SEQUENCE, 2 unordered pieces.
ACCESSION AC105414
VERSION AC105414.4 GI:18642912
KEYWORDS HTG: HTGS_PHASEI; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 175645)
Waterston,R.H.
Direct Submission
Submitted (04-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Feb 9, 2002 this sequence version replaced gi:18483567.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
Contact: submission@wustl.edu
----- Project Information -----
Center project name: HN0219L11
----- Summary Statistics -----
Sequencing vector: M13; 0%
Chemistry: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 174714 bases at least Q40
Consensus quality: 174877 bases at least Q30
Consensus quality: 174989 bases at least Q20
Insert size: 177000; agarose-fp
Insert coverage: 175545; sum-of-contigs
Quality coverage: 10.91 in Q20 bases; agarose-fp
Quality coverage: 8.85 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* *
* * 74211: contig of 74211 bp in length
* 74212 74311: gap of unknown length
* 74312 175645: contig of 101334 bp in length.
FEATURES             source
       source
           1..175645
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosome="4"
               /clone="RP11-219L11"
misc_feature         1..74211
                       /note="assembly_name:Contig34"
misc_feature         74312..175645
                       /note="assembly_name:Contig35"
BASE COUNT          54056 a 32068 c 32415 g 56997 t   109 others
ORIGIN

Query Match              27.8%; Score 630; DB 2; Length 175645;
Best Local Similarity    99.9%; Pred. No. 0;
Matches 680; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 44 ttcttttcttggtgggaactgcttagaggagggaggagggagaaagtgaatgaaa 103

```

Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)  
NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).

FEATURES  
source Location/Qualifiers  
i. 1871  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="LNG08487"  
/tissue\_type="human lung"  
/clone\_lib="LNG"  
/note="cloning vector pME18SFL3"  
204..713  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="BAB15689.1"  
/db\_xref="GI:10440272"  
/translation="MAPQQAWGSLQLSSQSECTGRLLTGNPGYSDWVPQSFNS  
LGNRNRNERITLSTGKSWLFLNITADPYERVDLSNRPGIVKLLRLSOFNKTAVP  
VRYPPKDRNSPRLNGVWGPWYKEETKKKPKSKNOAEKKQKKKKKKQKRAVSGS  
TCHSGVTCG"  
BASE COUNT 604 a 367 c 354 g 546 t  
ORIGIN

Query Match 11.6%; Score 263; DB 9; Length 1871;  
Best Local Similarity 99.1%; Pred. No. 9.3e-143;  
Matches 753; Conservative 0; Mismatches 5; Indels 2; Gaps 2;  
QY 1495 agattgatgaggaacattcaacatagatggtctatgatctctggagaccataagttagggtc 1554  
|||||  
Db 84 AGATTGATGAGGACATTCAACTAGATGCTATGATCTGGGAGACCATAAGTGAGGGTC 143  
QY 1555 ttgcctcaccctggagcgaacccggtggcacaatgaacccatatacacaagcacaataa 1614  
|||||  
Db 144 TTCGCTCACCCTGGAGTAGATATTTGGTACATGACCTTGACCCCATATACACCAAGCAAAA 203  
QY 1615 atggtcctctggcagcaggtctatgggtctggaacactgcaatcagtcagccatcagag 1674  
|||||  
Db 204 ATGGCTCCTGGCAGCAGGCTATGGGATCTGGAACACTTGCAATCCAGTCAGCCATCAGAG 263  
QY 1675 tgcagcact-ggaattgcttaaggaaatccttggtctacagcagctgggtccccctcag 1733  
|||||  
Db 264 TGCAGCACTGGGAGATTGCTTTACAGGAAATCCTGGCTACACGACTGGGTGCCCCCTCAG 323  
QY 1734 tcttccagcaacctgggacccaacccggtggcacaatgaacggatcacctcgtcaactggc 1793  
|||||  
Db 324 TCTTTCAGCACTGGGACCGAACCGGTGGCGCAATGACCGATCACCTTCTCAACTGGC 383  
QY 1794 aaaagtgtatggtctttcaacatcacagccagccatataatgagaggttggaacctatcaac 1853  
|||||  
Db 384 AAAAGTGTATGGCTTTTCAACATCACAGCCGACCATATGATGAGGGTGGACCTATCTAAC 443  
QY 1854 aggtatccaggaactcgtgaagaactcctcagggggtctctcacagttcaacaaactgca 1913  
|||||  
Db 444 AGGTATCAGGAATCGTGAAGAAGCTCCTACGGAGGCTCTCACAGTTTCAACAAACTGCA 503  
QY 1914 gtccggtccaggtatcccccaagaccccaagtagtaaccttagctcaatgaggaggtc 1973  
|||||  
Db 504 GTGCCGCTCAGGTATCCCCCAAGACCCCAAGAGTAGTACCCCTAGCTCAATGGAGGGGTC 563  
QY 1974 tagggacatggtatagagagggaaccaaagaaaagcgaagcacaagaaatcaggctgag 2033  
|||||  
Db 564 TGGGACCATGTTATAAGAGAGAAACCAAGAAAAGAGCCACGACAAATCAGGCTGAG 623  
QY 2034 aaaaagcaaaagaaagc-aaaaaagagaagaaacagcagaagcagaagcagtcagggttc 2092  
|||||

Db 624 AAAAAAGCAAAAGAAAAAGCAAAAGAAAAAGAAAGAAACAGCAAGACAGTCTCAGGTTCA 683  
QY 2093 acttgccattcaggtgttaacttggtgataagcacaataatattcttggtaaacttt 2152  
|||||  
Db 684 ACTTGCCCAATTGAGGTGTTACTGTGGATAAGCACAATAATTTCCTGTTGGTTAAACFTT 743  
QY 2153 aatcagttcttatcttctatctatctgttcttccctagtagtaaacacagcaaaatttgctcgataaat 2212  
|||||  
Db 744 AATCAGTCTTATCTTTCATCTGTTTCTTAGTAGTAACCAAGCAAAATTTGGCTCGATAATAT 803  
QY 2213 cgctggcctaagcgtcaggctgttttctatcgtgtgccac 2252  
|||||  
Db 804 CGCTGGCCTAAGCGTCAGGCTTGTTCATGCTGTGCCAC 843  
RESULT 9  
AC105383  
LOCUS  
DEFINITION Homo sapiens chromosome 4 clone RP11-9G1, WORKING DRAFT SEQUENCE, 4  
unordered pieces.  
AC105383 AC031992  
VERSION AC105383.2 GI:18376928  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 176341)  
AUTHORS Waterston,R.H.  
TITLE The sequence of Homo sapiens clone  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 176341)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (03-JAN-2002) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
COMMENT On Jan 26, 2002 this sequence version replaced gi:18042331.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc/index.shtml>  
Contact: [submissions@wustl.wustl.edu](mailto:submissions@wustl.wustl.edu)  
----- Project Information -----  
Center project name: H\_NH0009G01  
Drafting center: WIBR  
----- Summary Statistics -----  
Sequencing vector: M13; 34%  
Sequencing vector: plasmid; 66%  
Chemistry: Dye-primer ET; 0% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 173694 bases at least Q40  
Consensus quality: 174243 bases at least Q30  
Consensus quality: 174882 bases at least Q20  
Insert size: 180000; agarose-fp  
Insert size: 176041; sum-of-contigs  
Quality coverage: 11.08 in Q20 bases; agarose-fp  
Quality coverage: 10.78 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 1053: contig of 1053 bp in length  
\* 1054 1153: gap of unknown length  
\* 1154 5581: contig of 4428 bp in length



TITLE  
JOURNAL

## COMMENT

```

/note="assembly_fragment
clone_end:T7
vector_side:right"
BASE COUNT 70795 a 45474 c 46842 g 69434 t 406 others
ORIGIN

Query Match      2.3%; Score 53; DB 2; Length 232951;
Best Local Similarity 100.0%; Pred. No. 7e-19;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1611 aaaaatggtctctggcagcgatggtggtatctgggaactcgaatccagtc 1663
|||||
Db 146855 AAAAAAGCTCTGGCGAGCGATGATGGATCTGGACACATGCAATCCAGTC 146803

RESULT 11
AC098604 AC098604 129992 bp DNA linear HTG 20-DEC-2001
LOCUS Rattus norvegicus clone CH230-139G21, *** SEQUENCING IN PROGRESS
DEFINITION ***, 50 unordered pieces.
ACCESSION AC098604
VERSION AC098604.4 GI:17973771
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 129992)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alisbrooks,S.L., Amaralunga,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogue,M., Holloway,C.,
Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegh,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., Mcleod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenkwo,S.,
Ogah,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wlecyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

TITLE
JOURNAL
REFERENCE
2 (bases 1 to 129992)
Worley,K.C.

```

```

Direct Submission
Submitted (26-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:17064604.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GKIS
Center clone name: CH230-139G21
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 115001 bases at least Q40
Consensus quality: 121003 bases at least Q30
Consensus quality: 125541 bases at least Q20
Estimated insert size: 103320; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 1.9x in Q20 bases; sum-of-contigs estimation
-----

```

```

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 50 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

```

```

1 5288: contig of 5288 bp in length
5289 5388: gap of unknown length
5389 12822: contig of 7434 bp in length
12833 12922: gap of unknown length
12923 18920: contig of 5998 bp in length
18921 19020: gap of unknown length
19021 24271: contig of 5251 bp in length
24272 24371: gap of unknown length
24372 29092: contig of 4721 bp in length
29093 29192: gap of unknown length
29193 33320: contig of 4128 bp in length
33321 33420: gap of unknown length
33421 38234: contig of 4814 bp in length
38235 38334: gap of unknown length
38335 41472: contig of 3138 bp in length
41473 41572: gap of unknown length
41573 44757: contig of 3184 bp in length
44757 44856: gap of unknown length
44857 49504: contig of 4648 bp in length
49505 49604: gap of unknown length
49605 51767: contig of 2163 bp in length
51768 51867: gap of unknown length
51868 53489: contig of 1622 bp in length
53490 53589: gap of unknown length
53590 54945: contig of 1356 bp in length
54946 55045: gap of unknown length
55046 56965: contig of 1920 bp in length
56966 57066: gap of unknown length
57067 60629: contig of 3564 bp in length
60630 60729: gap of unknown length
60730 63415: contig of 2686 bp in length
63416 63515: gap of unknown length
63516 65243: contig of 1727 bp in length
65243 65342: gap of unknown length
65343 67014: contig of 1672 bp in length
67015 67114: gap of unknown length
67115 68525: contig of 1411 bp in length
68526 68625: gap of unknown length
68626 70727: contig of 2102 bp in length
70728 70827: gap of unknown length

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS

```

\* 70828 73751: contig of 2924 bp in length  
\* 73752 73851: gap of unknown length  
\* 73852 76742: contig of 2891 bp in length  
\* 76743 76842: gap of unknown length  
\* 76843 76883: contig of 1841 bp in length  
\* 76883 78784: gap of unknown length  
\* 78784 80648: contig of 1865 bp in length  
\* 80648 80748: gap of unknown length  
\* 80748 81958: contig of 1210 bp in length  
\* 81958 82058: gap of unknown length  
\* 82058 83786: contig of 1728 bp in length  
\* 83786 83887: gap of unknown length  
\* 83887 85421: contig of 1535 bp in length  
\* 85421 85522: gap of unknown length  
\* 85522 87315: contig of 1794 bp in length  
\* 87315 87415: gap of unknown length  
\* 87415 89651: contig of 2236 bp in length  
\* 89651 89752: gap of unknown length  
\* 89752 91450: contig of 1699 bp in length  
\* 91450 91550: gap of unknown length  
\* 91550 93240: contig of 1690 bp in length  
\* 93240 93341: gap of unknown length  
\* 93341 95063: contig of 1725 bp in length  
\* 95063 95163: gap of unknown length  
\* 95163 96780: contig of 1615 bp in length  
\* 96780 96880: gap of unknown length  
\* 96880 98469: contig of 1589 bp in length  
\* 98469 98569: gap of unknown length  
\* 98569 99694: contig of 1125 bp in length  
\* 99694 99794: gap of unknown length  
\* 99794 101631: contig of 1837 bp in length  
\* 101631 101732: gap of unknown length  
\* 101732 103235: contig of 1504 bp in length  
\* 103235 103335: gap of unknown length  
\* 103335 104663: contig of 1328 bp in length  
\* 104663 104763: gap of unknown length  
\* 104763 107050: contig of 2287 bp in length  
\* 107050 107150: gap of unknown length  
\* 107150 108692: contig of 1542 bp in length  
\* 108692 108792: gap of unknown length  
\* 108792 109912: contig of 1120 bp in length  
\* 109912 110012: gap of unknown length  
\* 110012 111516: contig of 1504 bp in length  
\* 111516 111616: gap of unknown length  
\* 111616 113034: contig of 1418 bp in length  
\* 113034 113134: gap of unknown length  
\* 113134 114537: contig of 1403 bp in length  
\* 114537 114637: gap of unknown length  
\* 114637 116034: contig of 1397 bp in length  
\* 116034 116134: gap of unknown length  
\* 116134 117734: contig of 1600 bp in length  
\* 117734 117834: gap of unknown length  
\* 117834 119020: contig of 1186 bp in length  
\* 119020 119120: gap of unknown length  
\* 119120 120235: contig of 1115 bp in length  
\* 120235 120335: gap of unknown length  
\* 120335 121616: contig of 1281 bp in length  
\* 121616 121716: gap of unknown length  
\* 121716 122992: contig of 1276 bp in length.  
FEATURES  
source  
1..122992  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="CH230-139G21"

BASE COUNT 35212 a 23990 c 23255 g 35545 t 4990 others  
ORIGIN

Query Match 1.3%; Score 29; DB 2; Length 122992;  
Best Local Similarity 100.0%; Pred. No. 0.0001;  
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 663 gagaaactactatgtccaccctatttgcac 691

||||| GAGAACTACTATGTCCACCTATTTCAC 119505  
Db 119477  
RESULT 12  
BC009528  
LOCUS Homo sapiens, clone IMAGE:3890809, mRNA, partial cds.  
DEFINITION BC009528  
ACCESSION BC009528.1 GI:16306920  
VERSION  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 927)  
Direct Submission  
AUTHORS Strausberg, R.  
TITLE  
JOURNAL  
REMARK  
COMMENT  
NIH-MGC Project URL: http://mgc.nci.nih.gov  
Contact: MGC help desk  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: DCTD/BTP/Gazdar  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: http://www.shgc.stanford.edu  
Contact: (Dickson, Mark) mcd@paxil.stanford.edu  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
Series: IRAC Plate: 14 Row: e Column: 7  
This clone was selected for full length sequencing because it  
passed the following selection criteria: Hexamer frequency ORF  
analysis.  
FEATURES  
source  
1..927  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3890809"  
/tissue\_type="Lung, carcinoma, large cell  
undifferentiated."  
/clone\_lib="NIH-MGC\_69"  
/lab\_host="DH10B"  
/note="vector: pCMV-SPORT6"  
<1..194  
/codon\_start=3  
/product="Unknown (protein for IMAGE:3890809)"  
/protein\_id="AAH09528.1"  
/db\_xref="GI:16306921"  
/translation="DASAHAHASVADIPGPHAAWQISEKWFNSCGAVCLISPGSL  
PLTSKGSLSCLRLVSYMGL"  
BASE COUNT 267 a 189 c 189 g 282 t  
ORIGIN

Query Match 1.1%; Score 26; DB 9; Length 927;  
Best Local Similarity 100.0%; Pred. No. 0.0069;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 caccgctccgccaccgctccgtgga 26  
Db 15 CACGGCTCCGCCACGCGTCCGTGGA 40

RESULT 13  
BC012659



```

LOCUS      BC012659      2003 bp      mRNA      linear      ROD 20-AUG-2001
DEFINITION Mus musculus, Similar to COP9 (constitutive photomorphogenic),
             subunit 7b (Arabidopsis), clone MGC:13817 IMAGE:4017012, mRNA,
             complete cds.
ACCESSION   BC012659
VERSION     BC012659.1 GI:15215084
KEYWORDS    MGC.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            Strausberg, R.
REFERENCE   1 (bases 1 to 2003)
AUTHORS    Direct Submission
TITLE      Submitted (15-AUG-2001) National Institutes of Health, Mammalian
JOURNAL    Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK     NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-re@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbio.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 18 Row: f Column: 6
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 3309175.

FEATURES             Location/Qualifiers
     source           1..2003
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="MGC:13817 IMAGE:4017012"
                     /tissue_type="Mammary tumor metastasized to lung. Tumor
                     arose spontaneously from a senescent normal mammary
                     (clonal) outgrowth infected with the virus MMV."
                     /clone_lib="NCI-CGAP-Lu29"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
                     /codon_start=1
                     /product="Similar to COP9 (constitutive photomorphogenic),
                     subunit 7b (Arabidopsis)"
                     /protein_id="AAH12659.1"
                     /db_xref="GI:15215085"
                     /translation="MAGEQKPPSNLLPQFTLLAKGTSGSALTTLISQVLEAPGVYVFG
                     ELLELANVQELAGANAYLQLNLFAYGYTPDYIANKESLPELSAQQNKHLITV
                     SLARMKIPYVLLKDLKMRNLELLEIIAYYTDIIQGLKDRNOLLEVDFCIGR
                     DIRKKDINNTVKTLEDCGEAVLLEIGIEQVLRANQYKKNHRTQOAEVSNIKK
                     TLKATASSAQAEQQLAERECPHPTEQROPTRKMSKVGLVSSRH"

BASE COUNT  465 a 541 c 516 g 481 t
ORIGIN

Query Match      1.1%, Score 26; DB 10; Length 2003;
Best Local Similarity 100.0%; Pred. No. 0.0067;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1  caccgcgtccgccacgcgtccgtgga 26
    ||||||
Db   2  CACGCGTCCGCCACGCGTCCGTGGA 27

RESULT  14
AC108339/c
LOCUS      AC108339      28735 bp      DNA      linear      HTG 27-JAN-2002

```

```

DEFINITION Rattus norvegicus clone CH230-109H4, *** SEQUENCING IN PROGRESS
            ***, 26 unordered pieces.
ACCESSION   AC108339
VERSION     AC108339.1 GI:18377119
KEYWORDS    HTG: HTGS_PHASE1.
SOURCE      Norway rat.
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 28735)
AUTHORS    Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
            Ausbrooks, S.L., Amaratunga, H.C., Are, J.R., Ayale, M., Banks, T.,
            Barbara, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,
            Bouck, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,
            Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
            Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
            Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
            Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
            Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
            Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
            Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
            Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
            Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
            Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
            Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
            Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
            Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
            Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
            Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
            Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
            Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
            Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulsegad, H.,
            Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
            Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
            Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
            Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
            Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
            Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,
            Oraguine, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
            Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
            Rives, M., Rojas, A., Rojebokan, I., Rolfe, M., Ruiz, S., Savery, G.,
            Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,
            Sodergren, E., Sonaika, T., Sparks, A., Stanley, H., Stone, H.,
            Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
            Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
            Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R.,
            Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, R.,
            Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
            Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
            Weinstein, G. and Gibbs, R.
            Direct Submission
            Unpublished
            2 (bases 1 to 28735)
            Worley, K.C.
            Direct Submission
            Submitted (27-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            ----- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ----- Project Information
            Center project name: GPM2
            Center clone name: CH230-109H4
            ----- Summary Statistics
            Sequencing vector: Plasmid; M77789
            Chemistry: Dye-terminator Big Dye; 98% of reads
            Assembly program: Phrap; version 0.990329First call to
            findPhrapList
            Consensus quality: 14825 bases at least Q40

```

Consensus quality: 18227 bases at least Q30  
Consensus quality: 20289 bases at least Q20  
Estimated insert size: 9365; sum-of-contigs estimation  
Quality coverage: 0x in Q20 bases; agarose-fp estimation  
Quality coverage: 0.1x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
consists of 26 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 884: contig of 884 bp in length  
885 984: gap of unknown length  
886 1846: contig of 862 bp in length  
887 1946: gap of unknown length  
888 1947: contig of 856 bp in length  
889 2802: gap of unknown length  
890 2803: contig of 783 bp in length  
891 3686: gap of unknown length  
892 3786: contig of 979 bp in length  
893 4765: gap of unknown length  
894 4865: contig of 1058 bp in length  
895 5922: gap of unknown length  
896 6022: gap of unknown length  
897 6936: contig of 914 bp in length  
898 7036: gap of unknown length  
899 7918: contig of 882 bp in length  
900 8018: gap of unknown length  
901 8475: contig of 457 bp in length  
902 8575: gap of unknown length  
903 9396: contig of 821 bp in length  
904 9496: gap of unknown length  
905 9497: contig of 874 bp in length  
906 10370: contig of 874 bp in length  
907 10371: gap of unknown length  
908 10471: contig of 1028 bp in length  
909 11498: gap of unknown length  
910 11599: contig of 908 bp in length  
911 12507: gap of unknown length  
912 12607: contig of 409 bp in length  
913 13016: gap of unknown length  
914 13116: gap of unknown length  
915 13933: contig of 818 bp in length  
916 14033: gap of unknown length  
917 14034: contig of 828 bp in length  
918 14862: gap of unknown length  
919 14962: contig of 1163 bp in length  
920 16125: gap of unknown length  
921 16225: contig of 690 bp in length  
922 16914: gap of unknown length  
923 17014: gap of unknown length  
924 17015: contig of 1124 bp in length  
925 18139: gap of unknown length  
926 18239: contig of 1177 bp in length  
927 19416: gap of unknown length  
928 19515: gap of unknown length  
929 21083: contig of 1568 bp in length  
930 21516: gap of unknown length  
931 21884: contig of 1756 bp in length  
932 22940: gap of unknown length  
933 23039: contig of 771 bp in length  
934 23810: gap of unknown length  
935 23811: contig of 1323 bp in length  
936 23911: gap of unknown length  
937 25233: gap of unknown length  
938 25234: contig of 1678 bp in length  
939 25334: gap of unknown length  
940 27012: contig of 1624 bp in length  
941 27112: Location/Qualifiers  
1. .28735  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"

BASE COUNT 8716 a 4798 c 5034 g 7614 t 2573 others

ORIGIN

Query Match 1.1%; Score 25; DB 2; Length 28735;  
Best Local Similarity 100.0%; Pred. No. 0.025;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 31 ttaacttttttttttttttttttttc 55  
Db 26740 TTAACCTTTTCTTTTCTTTTTC 26716

RESULT 15

CBRG36E21/c CBRG36E21 38866 bp DNA linear INV 04-NOV-2000  
LOCUS Caenorhabditis briggsae cosmid G36E21, complete sequence.  
DEFINITION AC084563  
ACCESSION AC084563  
VERSION AC084563.1 GI:11095013  
KEYWORDS HTG.  
SOURCE Caenorhabditis briggsae.  
ORGANISM Caenorhabditis briggsae  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;  
Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
REFERENCE 1 (bases 1 to 38866)  
AUTHORS Washington University Genome Sequencing Center.  
TITLE The C. briggsae Genome Sequencing Project  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 38866)  
AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (04-NOV-2000) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
COMMENT Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University,  
St. Louis, MO 63110, USA  
e-mail: jsplie@watson.wustl.edu

FEATURES  
source

1. 38866  
/organism="Caenorhabditis briggsae"  
/strain="GujArat G16"  
/db\_xref="taxon:6238"  
/clone="G36E21"

BASE COUNT 12988 a 6629 c 7483 g 11766 t  
ORIGIN

Query Match 1.1%; Score 25; DB 3; Length 38866;  
Best Local Similarity 100.0%; Pred. No. 0.024;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2043 aagaaaagcaaaaaaagaagaaga 2067  
Db 9383 AAGAAAAGCAAAAAAAGAAGAA 9359

Search completed: July 23, 2002, 19:09:16  
Job time: 14232 sec

FEATURES  
source

1. .28735  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"





ed Jul 24 10:05:49 2002

1789	tagatgaagcaatcaacaacgtagacattggtctaaagacttatggtttctataacaaca	1848	XX	Homo sapiens.
1282	gcattatctactcttcagataatggtgcccagctctacgagggagtaactggc	1341	XX	WO9946281-A2.
1849	gcattatctactcttcagataatggtgcccagctctacgagggagtaactggc	1908	PD	16-SEP-1999.
1342	ctctcagagtagcaaaagacatattggaaggaggtccgggctgaggtttgtgc	1401	PF	08-MAR-1999;
1909	ctctcagagtagcaaaagacatattggaaggaggtccgggctgaggtttgtgc	1968	XX	10-MAR-1998;
1402	atagcccaactctgaaaaaaggaacagtggtgaaggaaacctgtgcacatcactgact	1461	PR	98US-0077632.
1969	atagcccaactctgaaaaaaggaacagtggtgaaggaaacctgtgcacatcactgact	1521	PR	98US-0077641.
1462	ggtaccacactctcttctcactggtctgaagacagatgtagagacattcaactagatg	2088	PR	98US-0077649.
2029	ggtaccacactctcttctcactggtctgaagacagatgtagagacattcaactagatg	2088	PR	98US-0077791.
1522	gctatgatctggagacacataaagtgggtcttcctgcctcaccgagtagatatttgc	2148	PR	98US-0078004.
2089	gctatgatctggagacacataaagtgggtcttcctgcctcaccgagtagatatttgc	2148	PR	98US-0040220.
1582	ataacattgaccctatatacacaaggaacaaatggtcctgagcagaggtatggga	1641	PR	98US-0078866.
2149	ataacattgaccctatatacacaaggaacaaatggtcctgagcagaggtatggga	1701	PR	98US-0078910.
1642	tctggaacactgcaatccagtcagcagcagcagcagcagcagcagcagcagcagc	2267	PR	98US-0078936.
2208	tctggaacactgcaatccagtcagcagcagcagcagcagcagcagcagcagcagc	1761	PR	98US-0078939.
1702	atcctgggtacagagcagcagcagcagcagcagcagcagcagcagcagcagcagc	2327	PR	98US-0079294.
2268	atcctgggtacagagcagcagcagcagcagcagcagcagcagcagcagcagcagc	1821	PR	98US-0079656.
1762	ggcacaatgaacggtacacactgctcaactggtgcaaaagtgtatggctttcaacatcacag	2387	PR	98US-0079664.
2328	ggcacaatgaacggtacacactgctcaactggtgcaaaagtgtatggctttcaacatcacag	1881	PR	98US-0079689.
1822	ccagccatgagaggtgagcagcagcagcagcagcagcagcagcagcagcagcagc	2447	PR	98US-0079728.
2388	ccagccatgagaggtgagcagcagcagcagcagcagcagcagcagcagcagcagc	1941	PR	98US-0079786.
1882	tacgagagctctcagctcaacaaactgagcagcagcagcagcagcagcagcagcagc	2507	PR	98US-0079820.
2448	tacgagagctctcagctcaacaaactgagcagcagcagcagcagcagcagcagcagc	2001	PR	98US-0079923.
1942	ccagaagtaacctgagcagcagcagcagcagcagcagcagcagcagcagcagcagc	2567	PR	98US-0080105.
2508	ccagaagtaacctgagcagcagcagcagcagcagcagcagcagcagcagcagcagc	2060	PR	98US-0080107.
2002	agaaaaagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	2627	PR	98US-0080194.
2568	agaaaaagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	2107	PR	98US-0080327.
2061	agaagaaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc	2674	PR	98US-0080328.
2628	agaagaaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc		PR	98US-0080333.
RESULT 4				98US-0080334.
AAZ333987 standard; cDNA: 4650 BP.				98US-0081049.
ID	AAZ333987			98US-0081070.
XX				98US-0081071.
AC	AAZ333987;			98US-0081195.
XX				98US-0081203.
DT	07-DEC-1999 (first entry)			98US-0081229.
DE	Human PRO708 nucleotide sequence.			98US-0081817.
XX				98US-0081838.
XX	Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;			98US-0081952.
KW	probe; blood coagulation disorder; cancer; cellular adhesion disorder;			98US-0081955.
KW	secreted protein; transmembrane protein; ss.			98US-0082568.









Db 1909 ctctcagagcagcaagaaacatactggaagagaggtatccgggttaggtctgtgc 1968  
 QY 1402 atagccactctgaaacaaaggaacagtggtgaagaaacctgtgacacatcactgact 1461  
 Db 1969 atagccactctgaaacaaaggaacagtggtgaagaaacctgtgacacatcactgact 2028  
 QY 1462 ggtacccactctcattctcactggtgagagacagattgagagacattcaactagatg 1521  
 Db 2029 ggtacccactctcattctcactggtgagagacagattgagagacattcaactagatg 2088  
 QY 1522 gctatgatctggaagacataagtgaggtcttgcctcaccacccagtagattttgc 1581  
 Db 2089 gctatgatctggaagacataagtgaggtcttgcctcaccacccagtagattttgc 2148  
 QY 1582 ataacattgaccccatatcaccaaggaacaaatggtcctctggcagaggtatggga 1641  
 Db 2149 ataacattgaccc-tatacaccaaggaacaaatggtcctctggcagaggtatggga 2207  
 QY 1642 tctggaacactgcaatccagtcagccatcagagtcagcactggaattgttaccaggaa 1701  
 Db 2208 tctggaacactgcaatccagtcagccatcagagtcagcactggaattgttaccaggaa 2267  
 QY 1702 atctggctacagcagactgggtcccccctcagcttttcagcaacactgggaccgacggt 1761  
 Db 2268 atctggctacagcagactgggtcccccctcagcttttcagcaacactgggaccgacggt 2327  
 QY 1762 ggcacatgaacggtatcacctctcaactggtgcaaaagtgtatgcttttcaacatcacag 1821  
 Db 2328 ggcacatgaacggtatcacctctcaactggtgcaaaagtgtatgcttttcaacatcacag 2387  
 QY 1822 ccgaccatgatgaggggtggaacctatctaacaggtatccaggaatcgtgaagaagctcc 1881  
 Db 2388 ccgaccatgatgaggggtggaacctatctaacaggtatccaggaatcgtgaagaagctcc 2447  
 QY 1882 tacggaggtctcacagttcaacaaactcagtcagtcaggtatcccccacaaagacc 1941  
 Db 2448 tacggaggtctcacagttcaacaaactcagtcagtcaggtatcccccacaaagacc 2507  
 QY 1942 ccgaagtaacctagctcaatggaagggtctgaggaacctatggtatagagagaaacca 2001  
 Db 2508 ccgaagtaacctagctcaatggaagggtctgaggaacctatggtatagagagaaacca 2567  
 QY 2002 agaaaaagagccagcaaaatcagctgagaaaaagcaaaagaaagc-aaaaaaga 2060  
 Db 2568 agaaaaagagccagcaaaatcagctgagaaaaagcaaaagaaagc-aaaaaaga 2627  
 QY 2061 agaaagaaagcagaaagcagctcaggttcaacttgcattcaggt 2107  
 Db 2628 agaaagaaagcagaaagcagctcaggttcaacttgcattcaggt 2674

## RESULT 5

AAH46867  
ID AAH46867 standard; cDNA; 1800 BP.

XX  
AC AC  
XX XX

DT DT  
XX XX

DE Human 23553 sulfatase polypeptide coding sequence.

XX  
KW Sulfatase; 23553 sulfatase; nootropic; neuroprotective; antibacterial;  
 KW antiinflammatory; vasotropic; antitumor; gene therapy; human; ss.  
 XX OS, Homo sapiens.

XX  
FH. Key  
FT CDS  
FT CDS  
FT CDS

XX  
FT. Location/Qualifiers  
 FT. 1..1800  
 FT. /\*tag= a  
 FT. /product= "23553 sulfatase"

XX  
PN W0200155411-A2.

XX 02-AUG-2001.  
 XX 31-JAN-2001; 2001WO-US03266.  
 XX 31-JAN-2000; 2000US-0495823.  
 XX (MILL-) MILLENNIUM PHARM INC.  
 XX Glucksmann MA, Williamson M, Rudolph-Owen LA, Tsai F;  
 XX WPI; 2001-476214/51.  
 XX P-PSDB; AAB85483.  
 XX Novel human sulfatase polypeptides useful for treating and diagnosing  
 PT sulfatase-related disorders such as cerebrovascular diseases, acute  
 PT meningitis, multiple sclerosis, degenerative diseases and tumor  
 XX Claim 2; Fig 15; 180pp; English.  
 XX The invention provides 22438, 23553, 25278 or 26212 human sulfatase  
 CC polypeptides and polynucleotides. The sulfatase genes and polypeptides  
 CC are useful for treating disorders involving the brain such as  
 CC cerebrovascular diseases, infections such as acute meningitis,  
 CC demyelinating diseases including multiple sclerosis, degenerative  
 CC diseases affecting the cerebral cortex including Alzheimer's disease  
 CC and Pick disease, spinocerebellar degenerations including spinocerebellar  
 CC ataxias including Friedreich ataxia, and ataxia telangiectasia,  
 CC degenerative diseases affecting motor neurons including amyotrophic  
 CC lateral sclerosis, inborn errors of metabolism such as leukodystrophies,  
 CC toxic and acquired metabolic diseases, including vitamin deficiencies,  
 CC and neurocutaneous syndromes (phakomatoses) including neurofibromatosis.  
 CC The present sequence represents a human 23553 sulfatase polypeptide  
 CC coding sequence.  
 XX Sequence 1800 BP; 534 A; 434 C; 423 G; 409 T; 0 other;

Query Match 78.6%; Score 1781.6; DB 22; Length 1800;  
 Best Local Similarity 99.7%; Pred. No. 0;  
 Matches 1795; Conservative 0; Mismatches 4; Indels 1; Gaps 1;  
 QY 324 atggctccaggggtgtgtggggcctccgctcccttccacagcctgtgtgtgt 383  
 Db 1 atggctccaggggtgtgtggggcctccgctcccttccacagcctgtgtgtgt 60  
 QY 384 cctggaaaagatgctagcaaatggggcgctggcaggattctggtatcctcctccact 443  
 Db 61 cctggaaaagatgctagcaaatggggcgctggcaggattctggtatcctcctccact 120  
 QY 444 tatggttacctgtcctggggcagccttagaagagagagaaagggccttactagct 503  
 Db 121 tatggttacctgtcctggggcagccttagaagagagagaaagggccttactagct 180  
 QY 504 caagctggagaaactagagccacagcaacttccactccacgcccactctcatttc 563  
 Db 181 caagctggagaaactagagccacagcaacttccactccacgcccactctcatttc 240  
 QY 564 atctagcggatgatacaggatttagagatgtgggtaccacggatctagataaaca 623  
 Db 241 atctagcggatgatacaggatttagagatgtgggtaccacggatctagataaaca 300  
 QY 624 cctactctgacaagctcgtccgagaggttaaactgaaactgagaaactatgtccagcct 683  
 Db 301 cctactctgacaagctcgtccgagaggttaaactgaaactgagaaactatgtccagcct 360  
 QY 684 attgcacaccatccaggagtagtttattactggaaagtatcagatacacaccggactt 743  
 Db 361 attgcacaccatccaggagtagtttattactggaaagtatcagatacacaccggactt 420  
 QY 744 caacattctatcatagacactaccacccaaactgtttacctctgagaaatgcacacta 803  
 Db 421 caacattctatcatagacactaccacccaaactgtttacctctgagaaatgcacacta 480

